Analysis and Forecasting of Agricultural Commodity Prices

ePMA Prototype

Vitor Miguel Marques de Matos Abreu

1 Instituto Superior Técnico

Abstract

Agricultural market participants are exposed to considerable risks from varying availability and price of agricultural commodities, which may significantly affect the economic viability of their activities. In developed economies, there is a growing use of contractual instruments that mitigate the trade uncertainty inherent to agricultural markets and contribute to the stabilization of the supply and demand equilibrium. In this context, predictive price modelling is an important tool to support trade and investment decisions of producers, distribution, agribusiness and financial institutions operating in market segments related with agriculture.

The ePMA prototype is an information system that performs an automated analysis of the evolution of agricultural commodity prices and global weather conditions, making price projections based on the available information and the relationship between the monitored variables. The prediction function is constructed by machine learning on the available data series, using support vector regression.

When compared with some prevailing forecasting methodologies, the ePMA prototype provides forecasts with lower mean absolute percent errors over the last stages of the crop’s cycle of wheat and soybean. The observed lower mean absolute percent errors over the last three months of the crop’s cycle results mainly, from the inclusion of climate data that anticipates, to some extent, overall northern hemisphere crop growing conditions. The ePMA prototype is essentially automated and operates on a relatively limited data set, making it potentially usable for regional crops and markets with limited available information.

Keywords

Agricultural commodities, futures contracts, support vector regression, corn, soybean, wheat.

1 Introduction

Agricultural market participants are exposed to considerable risks from varying availability and price of agricultural commodities, which may significantly affect the economic viability of their activities. In developed economies, there is a growing use of contractual instruments that mitigate the trade uncertainty inherent to agricultural markets and “according to estimates of the United States Department
of Agriculture, agricultural contracts covered 41 percent of the value of U.S. agricultural production in 2005, up from 39 percent in 2003, 36 percent in 2001, 28 percent in 1991, and 11 percent in 1969\textsuperscript{1}. Within this context, predictive price modelling is an important tool to support trade and investment decisions of producers, distribution, agribusiness and financial institutions operating in market segments related with agriculture. Forward agricultural commodity trading contracts may be agreed directly between the parties concerned or through standard contractual instruments as futures contracts, which set the price and delivery date of the commodities and are tradable on the futures market.

2 Literature review and general framework of predictive price modelling

According to the efficient market hypothesis [Fama 1970], there are two general and verifiable assumptions: prices reflect all publicly available information; and prices change efficiently to reflect new public information. The relative availability of trading opportunities in financial markets can be understood as a degree of market inefficiency [Aldridge, 2010]. Under open and competitive trading conditions in the financial markets, although some anomalies may occur, the price is not expected to move significantly and persistently away from the value of the underlying asset (speculative bubble) and the market is not expected to create significant and recurring opportunities to anticipate short-term price evolution (arbitrage).


In the article “Climate Trends and Global Crop Production Since 1980” [Lobell et al., 2011], it is found that in the cropping regions and growing seasons of most countries, with the important exception of the United States, temperature trends from 1980 to 2008 exceeded one standard deviation of historic year-to-year variability. From the authors’ perspective, efforts to model the effects of climate on prices or food availability, even for individual countries, must consider effects throughout the world, given that agricultural commodities are traded worldwide and that world market prices are determined by global supply and demand.

In an article considered to be fundamental in the field of Machine Learning and Statistics, “An Overview of Statistical Learning Theory” [Vapnik, 1999], Vladimir N. Vapnik, shows how the abstract learning theory established conditions for generalization which are more general than those discussed in classical statistical paradigms and how the understanding of these conditions inspired new algorithmic approaches to function estimation problems.

In analysing diverse predictive pricing modelling methodologies, it is observed that the statistical instruments and algorithms used are to some extent related to the purpose of the modelling itself. From an economic perspective, with broader forecasting horizons (from a few months to several years) and

where more structural adaptations and adjustments are sought, it is common to use a more explanatory and fundamental approach based on statistical analysis. On the other hand, in essentially financial perspectives, with shorter forecasting horizons (from a few seconds to a few months) and where essentially seeking to manage the valuation and risk of a portfolio of financial assets, there is generally a preference for machine learning methods, with reduced explanatory value.

To predict or to understand? It appears to be a paradox at various theoretical and application levels, as it is difficult to identify the relationships between the variables obtained by machine learning, despite their statistical significance and effectiveness.

3 Predictive modelling of agricultural commodity prices

3.1 General Concepts of Agricultural Commodity Pricing

Under normal market conditions, price is determined by supply and demand equilibrium. The Supply and Demand Law and its application in the fundamental analysis of the agricultural commodities market is based on the assumption of rational behaviour of producers and consumers.

Stocks to Use Ratio is an indicator of the supply and demand situation for agricultural commodities. The relationship between stocks and consumption indicates the level of stocks of a given commodity as a percentage of total demand or consumption, according to:

\[
 Stocks \ to \ Use \ Ratio = \frac{Beginning \ Stocks + Annual \ Production - Annual Use}{Annual \ Use} \tag{3.1}
\]

Beginning stocks represent the final stocks of the previous year. Annual Production represents the total production for a given year. Annual Use is the sum of all final consumption, including human consumption, exports, seeds and animal feed.

By comparing current year Stocks to Use Ratio with other years, an estimate can be made as to the direction of the price trend as well as the likely extent of price changes.

USDA’s monthly World Agricultural Supply and Demand Estimates Report (WASDE) is the main reference information about the situation of world agricultural commodity markets. The report consists essentially of statistical aggregates, estimates and projections of key production parameters for major crops and livestock productions on the main producing countries.

The methodology generally used to relate prices to ending stocks is based on equilibrium models for competitive markets with inventories. For annually produced commodities, such as corn and wheat, supply is a function of the previous year’s price. Demand is a function of prices in the current period and the previous year. Lagged prices are particularly important for crops used for livestock feeding, as livestock production decisions made in previous periods in response to prices in those periods affect

---

livestock inventories, and thus feed demand, over a number of years. Export demand would also be a function of lagged prices to reflect foreign supply response [Westcott et al, 1999].

USDA supply and demand estimates, included in the WASDE report, reflect a full balance sheet for each commodity and country. The process of forecasting price and balance sheet items is a complex one involving the interaction of expert judgment, commodity models, and in-depth research by USDA analysts on key U.S.A and international issues [Vogel et al, 1999].

3.2 ePMA prototype price forecasting methodology

The ePMA prototype is an information system that automatically analyses the evolution of agricultural commodity prices and global climate conditions, making projections based on the available information and the relationship between the monitored variables.

The prediction function is constructed by machine learning on the available data series, using support vector regression. The projection horizon covers the crop’s cycle from the start and is essentially aimed at anticipating the average farm price for agricultural commodities.

3.2.1 Support Vector Regression

Support Vector Machines are supervised learning models that analyse data for classification and regression.

In Support Vector Regression, given a training data set \{(x_1, y_1), \ldots, (x_l, y_l)\}, the goal is to find a function \(f(x)\) that has at most \(\varepsilon\) deviation from the actually obtained targets \(y_i\) for all the training data, and at the same time is as flat as possible.

According to Smola and Scholkopf [Smola et al, 2004], for the case of linear functions \(f\), taking the form,

\[
f(x) = \langle w, x \rangle + b \quad \text{with} \quad w \in X, b \in R
\]  

where \(\langle \cdot, \cdot \rangle\) denotes the dot product in \(X\), flatness in the case of (3.2) means that one seeks a small \(w\) and one way to ensure this is to minimize the norm, i.e. \(|w|^2 = \langle w, w \rangle\). This optimization problem can be stated, using slack variables \(\xi_i\) and \(\xi_i^*\), as:

- minimize\(\frac{1}{2}|w|^2 + C \sum_{i=1}^{l}(\xi_i + \xi_i^*)\) with \(C > 0\)  

\[
\text{subject to, } \begin{cases} 
\langle w, x_i \rangle - b - \varepsilon \leq \xi_i \\
\langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\
\xi_i, \xi_i^* \geq 0 \end{cases}
\]  

The constant \(C > 0\) determines the trade-off between the flatness of \(f\) and the amount up to which deviations larger than \(\varepsilon\) are tolerated. This corresponds to dealing with a so called \(\varepsilon\)-insensitive loss function \(|\cdot|_{\varepsilon}\) (3.5), where only the points outside the shaded region contribute to the cost insofar, as the deviations are penalized in a linear fashion,
In the context of the ePMA prototype and for the available data, it was observed that the projections obtained without data transformations had lower mean absolute percentage error, so that in the current implementation, a polynomial kernel parameterized for degree one is used.

In Support Vector Regression, the determination of the approximate function depends essentially on the atypical observations, which are at the margin or outside the zone between \( \varepsilon \) and \( -\varepsilon \) and therefore constitute the support vectors, that allows us to understand: model adjustment to situations determined by occurrence of relevant values only in some of the input variables (noise and outliers tolerance [Gama et al, 2017]); and how a large set of input variables with correlations that are not always significative, may be considered without resulting in erratic projections (robustness in the face of large objects [Gama et al, 2017]).

### 3.2.2 Economic data

For the economic analysis, the variables used to predict the evolution of agricultural commodity prices are:

- Price ratios between futures contracts with different delivery dates for the same commodity;
- Price ratios between futures contracts of different commodities.

Price ratios between futures contracts with different delivery dates for the same commodity are related to the stock-to-use ratio and largely reflect the level of stocks. In the context of the data used in the ePMA prototype, it is important to mention that futures contracts with the closest delivery date correspond to stock commodity prices and that futures contracts with the late delivery dates generally correspond to prices for the upcoming harvest.

The price ratios between different agricultural commodities used in animal feed, mainly reflect substitution effects in the livestock sector, as the various grains and oilseeds are largely interchangeable in animal feed and price is a key criterion.
3.3.3 Climate data

For the climate analysis, the variables used to predict the evolution of agricultural commodity prices are:

- Annual variation in southern hemisphere ocean surface temperature anomaly during spring and summer (trend of the year)
- Annual variation in northern hemisphere land surface temperature anomaly during spring and summer of previous year, compared to the average of the previous three years (trend of the negative feedback effects)

Abnormally high temperatures during the growth season of grains and oilseed crops generally have a negative effect on crop yield. The negative impact of the long-term trend of rising temperatures in key production regions is difficult to identify and quantify, since the development of agricultural technology has allowed continued productivity growth and the yield per unit area of the analysed crops has almost tripled over the past sixty years. However, there is some correlation (0.30) between the annual variation of the northern hemisphere temperature anomaly during spring and summer and the annual variation of the average farm price of agricultural commodities, with unexpected high temperatures having some impact on production and generally resulting in below average yields.

Estimation by multiple linear regression of the annual variation of the northern hemisphere land surface temperature anomaly during spring and summer, based on the annual variation of the southern hemisphere ocean surface temperature anomaly during spring and summer (trend of the year) and the variation of the northern hemisphere land surface temperature anomaly during spring and summer of the previous year compared to the average of the previous three years (trend of the negative feedback effects), shows a significant correlation (0.69) considering the complexity of the climate interactions considered.

![Figure 2](image.png)

**Figure 2:** Comparison between the observed (blue line) and the predicted (orange line) value for the annual variation of the land surface temperature anomaly in the northern hemisphere from April to September

4 ePMA prototype usage

The ePMA prototype implements a user interface to perform price projections and manage importation and consolidation of agricultural commodities futures data.
Consolidation of financial data consists mainly of two operations on the imported data: re-concatenation of futures contract; and obtaining additional data series by averaging the prices of the three closest to delivery futures contracts for each commodity.

5 Evaluation of results

There are several indicators and metrics for evaluating forecast errors. In this project, the evaluation indicator used was the mean absolute percent error of the predicted values, in order to obtain comparable results with relevant articles published about prevailing forecasting methodologies.

The performance of the ePMA prototype was verified against the performance of the WASDE report forecasts for the marketing years from 1980/1981 to 2006/2007, as documented by Isengildina-Massa, Irwin and Good in “Quantile Regression Estimates of Confidence Intervals for WASDE Price Forecasts” [Isengildina-Massa et al, 2010] and in relation to the performance of the predictive value of agricultural commodity futures, calculating the last price based on an average difference between the average farm price and the price of futures contracts (basis) of 9% of the futures contracts prices.

A 9% basis reflects a generic optimization of the predictive value of futures in determining the last price, for the analysed data series.

The ePMA prototype forecasts, have a mean absolute percent error intermediate between the WASDE report forecasts and the forecast that corresponds to using the last price as forecast (Table 1), except for corn where the last price as forecast has a lower mean absolute percent error.

<table>
<thead>
<tr>
<th>Forecasts (months)</th>
<th>ePMA</th>
<th>WASDE</th>
<th>Last price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn (May to November)</td>
<td>10.43</td>
<td>9.84</td>
<td>10.22</td>
</tr>
<tr>
<td>Wheat (May to August)</td>
<td>7.85</td>
<td>7.90</td>
<td>8.37</td>
</tr>
<tr>
<td>Soybean (May to November)</td>
<td>8.94</td>
<td>8.41</td>
<td>9.50</td>
</tr>
<tr>
<td>Average</td>
<td>9.07</td>
<td>8.72</td>
<td>9.36</td>
</tr>
</tbody>
</table>

Table 1 - Comparison of the average of the mean absolute percent error of the ePMA prototype price forecasts for the average farm price of corn, wheat and soybean with those published in the WASDE report and last price as forecast, from 1980/1981 to 2006/2007.

If we consider only the last three months of the crop cycle (Table 2), the ePMA prototype forecasts have an average mean absolute percent error lower than the WASDE report forecasts and the last price as forecasts, with the exception of corn where the last price as forecast has a lower error.

It can be verified using only subsets of the variables, that this performance results mainly from the inclusion of the climate variables, which according to the efficient market hypothesis, may correspond to information that the market players find more difficult to obtain and integrate in their decisions.


<table>
<thead>
<tr>
<th>Forecasts (months)</th>
<th>ePMA</th>
<th>WASDE</th>
<th>Last price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn (September to November)</td>
<td>7.04</td>
<td>7.34</td>
<td>6.96</td>
</tr>
<tr>
<td>Wheat (June to August)</td>
<td>6.66</td>
<td>7.17</td>
<td>7.44</td>
</tr>
<tr>
<td>Soybean (September to November)</td>
<td>5.61</td>
<td>6.52</td>
<td>6.77</td>
</tr>
<tr>
<td>Average</td>
<td>6.44</td>
<td>7.01</td>
<td>7.06</td>
</tr>
</tbody>
</table>

Table 2 - Comparison of the average of the mean absolute percent error of the ePMA prototype price forecasts for the average farm price of corn, wheat and soybean with those published in the WASDE report and last price as forecast, for the last three months of the crop’s cycle from 1980/1981 to 2006/2007.

When we consider only the last three months of the crop (Table 3) from 1980/1981 to 2017/2018, the ePMA prototype forecasts have a mean absolute percent error lower than last price as forecasts.


<table>
<thead>
<tr>
<th>Forecasts (months)</th>
<th>ePMA</th>
<th>Last price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn (September to November)</td>
<td>7.24</td>
<td>7.39</td>
</tr>
<tr>
<td>Wheat (June to August)</td>
<td>7.97</td>
<td>8.28</td>
</tr>
<tr>
<td>Soybean (September to November)</td>
<td>5.58</td>
<td>7.37</td>
</tr>
<tr>
<td>Average</td>
<td>6.93</td>
<td>7.68</td>
</tr>
</tbody>
</table>

Table 3 - Comparison of the average of the mean absolute percent error of the ePMA prototype price forecasts for the average farm price of corn, wheat and soybean with last price as forecast, for the last three months of the crops’ cycle from 1980/1981 to 2017/2018.
6 Conclusions

When compared with some prevailing forecasting methodologies, the ePMA prototype provides forecasts with lower mean absolute percent errors over the last stages of the crop’s cycle of wheat and soybean. The observed lower mean absolute percent errors over the last three months of the crop’s cycle results mainly, from the inclusion of climate data that anticipates, to some extent, overall northern hemisphere crop growing conditions.

Overall, the ePMA prototype has no advantage in the quality of price projections over the United States Department of Agriculture’s World Agricultural Supply and Demand Estimates (WASDE) report, because only forecasts for the last three months of the crop’s cycle present lower mean absolute percent error. If we consider the resources needed to prepare the WASDE report forecasts, which involve a comprehensive set of information and the involvement of a broad team of USDA staff and collaborators, there is some benefit as to the efficiency with which price projections are prepared. The ePMA prototype is essentially automated and operates on a relatively limited data set, making it potentially usable for regional crops and markets with limited available information.

Within the framework of agile methodologies, characterized by an iterative and incremental approach to software development, the ePMA prototype can be understood as an essential step towards the development of an agricultural market information system, which sought to delineate a generic methodology to forecast agricultural commodity prices and identify the information required for the operation of such a system.

Automated construction of prediction functions using support vector regression results generally in predictions with lower mean absolute percent error than the automated construction of prediction functions using multiple linear regression and the forecast derived from the last price of the futures contracts.
7 Bibliography

7.1 Books and Articles


7.2 Databases

