The influence of Brazilian crisis on inefficiency and nonlinearity on soybean prices

EVERTON ANGER CAVALHEIRO[1], KELMARA MENDES VIEIRA[2], ALISSON EDUARDO MAEHLER[3], CARLOS COSTA[4]
Centro de Engenharias
Federal University of Pelotas
Street Evandro Behr, 5992 apartament 302. City of Santa Maria. State of Rio Grande do Sul
BRAZIL
[1]eacavalheiro@hotmail.com, [2]kelmara@terra.com.br, [3]alisson.maehler@gmail.com, [4]carlos.costa1@gmail.com

Abstract: - We investigated the effect of the 2007-2008 Brazilian financial crisis on nonlinearity and the prediction accuracy of artificial neural networks on monthly soybean prices in Brazil. To determine the exogenous variable, the commodity's logarithm return was calculated. The best period for the series simulation was then identified, simulations carried out and the model validated. Model forecasting results were satisfactory for all samples. A Group Method of Data Handling (GMDH) methodology was capable of demonstrating the returns' non-randomness, denoting marketing inefficiency, arbitrage opportunities and abnormal return to investors, especially after the financial crisis of 2007-2008.

Key-Words: Financial Crisis, Predictability, Nonlinearity, Soybean

1 Introduction

Soybean [*Glycine max* (L.) Merr.], with its countless and varied uses, is an important crop at the global level. Its seeds are rich in oil - approximately 20% - and protein - approximately 40% - (Singh, 2010). Soybean is one of the oldest food sources known to humans. In 2011, the total cultivated area of soybean in the world was 102.99 million ha and the total production was 260 Tg y^{-1} (FAO, 2013), of which 75 Tg y^{-1} were produced in Brazil. Global soybean production and trade has changed dramatically in the past 30 years (Chianu *et al.*, 2010). These changes have been driven by an increasing demand for soybean meal, a component which accounts for 65% of animal feed bulk (Ash et al., 2006).

Growing economies such as those of China, India and other developing countries have dramatically increased the demand for livestock products, which, in turn, has increased the demand for soybean meal (Delgado, 1999). In 2007, the global area, production and productivity of soybean were 90.1 million ha, 220 Tg and 2.44 Mg ha^{-1}, respectively (Singh, 2010). According to this author, the USA, Brazil, Argentina, China and India are the major soybean-producing countries. IGC (2015) noted of that for a world soybean production of 316 Tg y^{-1} projected for 2014/2015, 265 Tg y^{-1} would come from the exporting countries of Brazil, Argentina and the USA, and the remainder from other countries.

Both worldwide and in Brazil, soybean is the crop having shown the greatest perceptual growth in last few years. According to USDA data , global soybean production grew from 44 Tg y^{-1} in 1970, to over 220 Tg y^{-1} in 2008. Considering the favourable weather conditions for the American harvest, August 2009 saw a projected (2010/2011) 15% increase (32 Tg y^{-1}) in global production, with a projected American production for 2010/2011 of 261 Tg y^{-1}, representing a four-fold growth. This represents substantial growth, compared to other crops: *e.g.*, 300 to 792 Tg y^{-1} (1.6-fold) of wheat (*Triticum æstivum* L.), and 310 to 432 Tg y^{-1} (40%) for rice (*Oryza sativa* L.), over the same period (Trennepohl & Paiva, 2011).

Accounting for 25.14% of global soybean production, in 2011 Brazil's produced 66 Tg y^{-1} of soybean on 2.5×10^8 km^2, and area

equivalent to all the UK territory (FAOSTAT, 2014). Still in 2010, soybean accounted for only 9% of all Brazilian exports, 5.6% of the nation's agricultural GDP and 1.25% of its overall GDP.

Many factors influence soybean prices (Jun and Chao, 2010), *e.g.*, meteorological conditions, the family consumption level, the consumption structure, offer and demand, as well as national and international stock in the futures market and the soybean circulation system. This circulation system has non-linear features typical of a dynamic system and of the evolution law. This is sustained by the application of the chaotic sequence in order to study fluctuation and price forecast law. All of these features influences the price efficiency.

Based on different observed references with respect to information type, Fama (1970) stated the efficient market hypothesis to be comprised of three forms: weak , semi-strong and strong. Weak form efficiency is based on a set of information that only includes price or stocks return history. The semi-strong form considers a set of information that only includes the public knowledge available to all participants in the market, while the strong form includes any information obtained by any participant in the market.

Other definitions of market efficiency have been suggested (Rubinstein, 1975; Jensen, 1978; Beaver, 1981; Black, 1986; Dacorogna *et al*., 2001; Malkiel, 2003; Timmermann and Granger, 2004; Milionis, 2007). Since there is no consensual definition for the pattern of market efficiency, we adopted the version enounced by Fama (1970) that emphasises both speed and precision of price adjustments to new information.

Interest in predicting the behavior of prices is probably as old as the markets themselves, so the literature on this matter is wide and significant. Recent studies, such as that of Righi & Ceretta (2011), implementing time series analysis, showed daily quotations for some Brazilian commodities [soybean, cotton (*Gossypium hirsutum* L.), coffee (*Coffea arabica* L.) and sweet corn (*Zea mays* L.)] do not follow anticipated market efficiency, thus generating opportunities arbitrage.

In Brazil, the study of efficiency and predictability in the agricultural commodity markets is important to government as well as producers and purchasers. For the government, an efficient market is a better alternative than market intervention through policies. For processors and marketers, predictability provides a reliable forecast of prices allowing them to effectively manage their market risks. It is also in the interests of international market participants from countries like Canada, the USA, Australia and the European Union, who are major grain exporters.

Considering these issues, we defined the following research problem: "Did the financial crisis of 2007 to 2008 influence the nonlinearity and prediction accuracy of soybean price paid in Brazil?"

2 Materials and Methods

In order to answer the research problem, we performed a time series (January 1990-May 2014) analysis using the logarithmic return of the prices paid to producers in Brazil as the exogenous variable. In order to calculate return, we used a secondary data base (IPEADATA, 2011).

Tsay (2005) states that two main reasons exist why most studies on financial time series use returns rather than assets themselves: (i) for the average investor, assets return is an adequate indicator when comparing investments opportunities and, (ii), return series are easier to deal with than a price series, since returns show more attractive statistical features, including the fact that non-bias is common in non-stationary data series. Given that, according to the assumed hypothesis, asset returns are independently and identically distributed (i.i.d.) with an average μ and variance σ^2, using logarithmic returns is well suited to financial studies (Tsay, 2005).

After calculating the logarithmic returns, we performed a random walk test to assess whether or not the data series presented non-random features. Results of this analysis indicated these features to be non-random, therefore offering the opportunity to perform modelling in an effort towards time series forecasting. In order to analyse the level of predictability of soybean price, we used the sample determination coefficient (R^2), which measures the proportion or percentage of variation in y anticipated by models:

$$R^2 = 1 - \frac{\sum_1^N (\hat{y}_i)^2}{\sum_1^N (y_i - \bar{y})^2}.$$

where, N is number of observations y_i is return during i period, $\hat{y}_i$ is computed values according to the model $\bar{y}$ is the mean value. Two other indicators were used, namely the mean standard error (MSE) and mean absolute error (MAE):

$$MSE = \sqrt{\frac{1}{N}\sum_1^N (y_i - \hat{y}_i)^2}.$$

$$MAE = \frac{1}{N}\sum_1^N \left|\sqrt{y_i^2} - \sqrt{\hat{y}_i^2}\right|.$$

We also analysed the Theil inequality coefficients, also known as U. The denominator for U is MSE, but the scale for the denominator is such that U exists in the interval from 0 to 1; where U=0 constitutes a perfect forecast of observed values, and U=1 the model's worst possible predictive performance:

$$U = \frac{\sqrt{\frac{1}{N}\sum_1^N (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N}\sum_1^N (y_i)^2} + \sqrt{\frac{1}{N}\sum_1^N (\hat{y}_i)^2}}.$$

Besides the Theil inequality coefficient, we analysed the bias proportion and variance proportion (U^M and U^S, respectively), allowing us to break down the error into its characteristics sources. The U^M addresses possible systematic error, since it measures how much the average values for the simulated and effective series deviate from each other (Pindyck and Rubinfield, 1991). Whatever the value of U, U^M is expected to be close to zero. An elevated value for U^M (above 0.1 or 0.2) would be worrying, since it would indicate the presence of systematic bias, requiring the model to be modified accordingly. The U^M and U^S are calculated as:

$$U^M = \frac{(\bar{y}^S - \bar{y}^A)^2}{(1/T)\sum(\bar{y}_t^S - \bar{y}_t^A)^2}.$$

$$U^S = \frac{(\sigma_S - \sigma_A)^2}{(1)^{(1/T)}\sum(\bar{y}_t^S - \bar{y}_t^A)^2}.$$

where, $\bar{y}^S, \bar{y}^A$ are the means of observed and estimated values, respectively, and σ_S and σ_A are the standard deviations of observed and estimated values, respectively. The variance proportion U^S, indicates the capacity to replicate the rate of variability rate of the variable of interest (Pindyck and Rubinfield, 1991). A high value of U^S would indicate that the effective series floated a great deal. That would also be worrying and could lead to revising the models. To further evaluate the forecasts' success the δ^2 was calculated (Ivakhnenko et al., 1993):

$$\delta_i^2 = \frac{\sum_1^N (y_i - \hat{y}_i)^2}{\sum_1^N (y_i - \bar{y})^2} \to \min.$$

Adequate performance would be reflected in cases where $\delta^2 \leq 0.05$, while a satisfactory performance would be reflected in cases where $0.5 < \delta^2 < 0.8$, while $\delta^2 > 1.0$ indicate inadequate performance and the need to revisit the modelling process.

To compare the accuracy of forecasts against random walk predictions, we used the Diebold-Mariano test (Diebold & Mariano, 1994) . When comparing two forecasts, the question of whether the predictions of a given model, A, are significantly more accurate, in terms of a loss function $g(\cdot)$, than those of the competing

model, *B* arises. The Diebold-Mariano test aims to test the null hypothesis of equality of expected forecast accuracy against the alternative of differing forecasting ability across models. The null hypothesis of the test can be, thus, written as:

$$d_t = E\left[g\left(e_t^A\right) - g\left(e_t^B\right)\right] = 0$$

where, e_t^i is the forecasting error of model *i* when performing *h*–step-ahead forecasts.

The Diebold-Mariano test uses the autocorrelation-corrected sample mean of d_t in order to test the null hypothesis (Eq. 8). If *n* observations and forecasts are available, the test statistic is, therefore,

$$S = \left[\hat{V}\left(\bar{d}\right)\right]^{-1/2}\bar{d}, \text{ where} \quad (9)$$

$$\hat{V}\left(\bar{d}\right) = \frac{1}{n}\left[\hat{\gamma}_0 + 2\sum_{k=1}^{h-1}\hat{\gamma}_k\right], \text{ and} \quad (10)$$

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=k+1}^{n}\left(d_t - \bar{d}_t\right)\left(d_{t-k} - \bar{d}_t\right) \quad (11)$$

Under the null hypothesis of equal forecasting accuracy, *S* is asymptotically normally distributed.

3. Results, Discussion, and Conclusions

Before analyzing predictability of Brazilian soybean return prices, it was necessary to choose the period to analyze. We therefore tried to identify the structural breaks in the analyzed series. The problem of detecting structural changes in linear relationships has been an important topic in econometric and statistical research (Zeileis et al., 2001), considering that a careless analysis can result in incorrect inferences in causality tests, co-integration and acceptance of incorrect models (Covas, 1997). The latter stated that these tests can determine the way in which exogenous shocks or political regime changes are felt in the behavior of some economic indicators.

In order to adequately treat the time series, some authors have presented several tests that make it possible to identify and estimate the moments for structural breaks. Among the first works to be published, we can find tests by Chow (1960) and cumulative sum control charts (CUSUM; Brown et al., 1975). The former tests had the inconvenience of implying an *a priori* knowledge of where the structural break was, while the latter test, part of a different class of tests, allows one to detect breaks of several types for parameters of interest, and for which one is not required to specify the number of breaks in the series (Covas, 1997).

In the present study we used a more sophisticated model to estimate the structural breaks in the data series. The Bai and Perron (1998) method allows one to simultaneously estimate multiple breaks as well as determine their previously unknown dates. Initially, we tested the hypothesis of the existence of structural breaks in the Brazilian soybean returns prices on a monthly basis. Based on the monthly negotiation prices for soybean, we calculated the logarithmic returns and rejected the null hypothesis stating that the vector *b* variance was constant throughout the whole series (*F* stats = 20.8434, sig. 0.000). This indicated the existence of structural breaks in the time series (Figure 1).

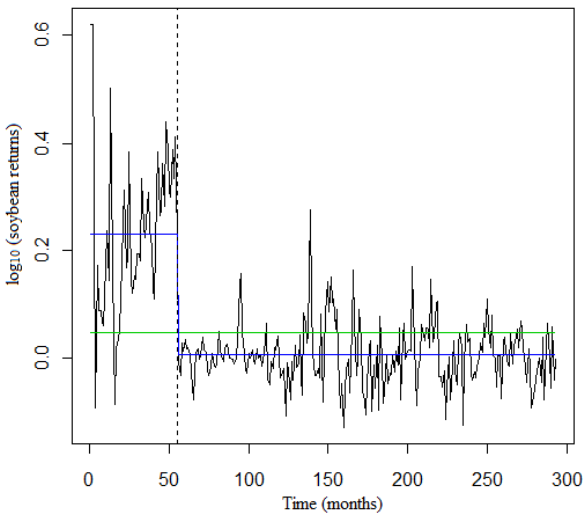

Figure 2: Structural Breaks on international soybean price´s

A single rupture in the monthly logarithmic returns of Brazilian soybean prices occurred between January 1990 and May 2014, as the change in behaviour of this series at the 55th observation attests (Figure 1). We therefore performed the simulation excluding the first 55 observations (January 1990 – June 1994).

Excluding the first 55 observations (January 1990 – June 1994) can be explained from an economics perspective: Brazil initiated the Plano Real, a program strictly limiting government spending, creating a new currency, and implementing many other fiscal reforms in June 1994. This significantly modified inflation memory within the Brazilian society, enhancing the Brazilian productive processes, especially the agribusiness sector and soybean trading.

A well-known problem in modelling is the search for an optimal model, which essentially depends on the adopted methodology. Prior to testing the predictability of the series, it was necessary to choose an appropriate method. The linearity test is a determinant criterion when choosing the methodology to be adopted when modelling a time series (Steyerberg, 2009). A similar problem occurs when one examines different transformations, jeopardizing the variable's linearity.

Several tests have been proposed for assessing the need for nonlinear modeling in time series analysis (Cryer and Chan, 2008). Some of these tests, such as those of Keenan (1985) and Tsay (1986), can be interpreted as Lagrange multiplier tests for specific nonlinear alternatives.

Keenan (1985) derived a test for nonlinearity analogous to Tukey's one degree of freedom for nonadditivity test. Keenan's test seeks to approximate a nonlinear stationary time series by a second-order Volterra expansion.

$$Y_t = \mu + \sum_{\mu=-\infty}^{\infty} \theta_\mu \varepsilon_{t-\mu} + \sum_{\nu=-\infty}^{\infty}\sum_{\mu=-\infty}^{\infty} \theta_{\mu\nu} \varepsilon_{t-\mu}\varepsilon_{t-\nu}$$

where,

Y_t is the process in time *t*, ε is the error of the process at the past (μ or ν).

In this case, $\{\varepsilon_t, -\infty<t<\infty\}$ is a sequence of independent and identically distributed zero-mean random variables. The process is linear if the double sum on the right hand side of Eq. 12 vanishes (Nazlioglu & Soytas, 2010). Keenan's test is equivalent to a test of n=0, according the regression model (Cryer and Chan, 2008):

$$Y_t = \theta_0 + 1 + \varphi_1 y_{t-1} + \ldots + \varphi_m y_{t-m} + \exp\left\{ n\left(\sum_{j=1}^{m} \varphi_j y_{t-j} \right)^2 \right\} + \varepsilon_t$$

where, Y_t is the process in time *t*, θ is a constant, $\varphi_1, \ldots, \varphi_m$ and η are the parameters of y_t at time *t*, and ε_t is the error at time *t*.

In this case, $\{\varepsilon_t\}$ are independent and normally distributed with zero mean and finite variance.

If $\eta \neq 0$, the model is non-linear. Keenan's test is both conceptually and computationally simple and only has one degree of freedom, which makes the test very useful for small samples (Cryer and Chan, 2008). However, Keenan's test is only powerful in detecting nonlinearity in the form of the square of the approximating linear conditional mean function. Tsay (1986) extended Keenan's approach by considering more general nonlinear alternatives. A more general alternative to nonlinearity may be formulated by replacing the term:

$$\exp\left\{ n\left(\sum_{j=1}^{m} \varphi_j y_{t-j} \right)^2 \right\} + \varepsilon_t$$

by

$$\begin{aligned} &\exp(\ \delta_{1,1} y_{t-1}^2 + \delta_{1,2} y_{t-1} y_{t-2} + \ldots + \delta_{1,m} y_{t-1} y_{t-m} \\ &+ (\delta_{2,2} y_{t-2}^2 + \delta_{2,3} y_{t-2} y_{t-3} + \ldots + \delta_{2,m} y_{t-2} y_{t-m} \ldots \\ &+ (\delta_{m-1,m-1} y_{t-m+1}^2 + \delta_{m-1,m} y_{t-m+1} y_{t-m} + \ldots + \delta_{m,m} y_{t-m} y_{t-m}) + \varepsilon \end{aligned} \quad (12)$$

where, ε_t is a white noise, φ_j and δ_m are the parameters at time *j* or *m*, for example.

Using the approximation exp(*x*) ≈ 1 + *x*, the nonlinear model can be approximated as a quadratic AR model, and whether or not all the *m*(*m* + 1)/2 coefficients $\delta_{i,j}$ are zero can be tested by an *F*-test (Cryer and Chan, 2008).

In order to test each regime's non-linearity, we used the Tsay Test (Tsay, 1986), which assesses the existence of non-linearity on average, and considers the residuals ($\hat{\varepsilon}_i$) of the auto-regressive process:

$$\hat{y}_i = \hat{\beta}_i y_{i-1} \cdots + \hat{\beta}_p y_{i-p+1} + \hat{\varepsilon}_i$$

where, $\hat{\varepsilon}_i$ represents the estimated residuals of the model, p is the number of lags, $\hat{y}_i$ is the estimated dependent variable, and y_{i-1} is the lagged dependent variable in *t*-1.

For each y_t observation, we built a vector z_t of the lagged variables' cross products, *i.e.*, y_{t-i} ,y_{t-j} for *i*, *j* = 1, ..., *p* where $i > j$. For instance, if $p = 2$ then $z_t = \left[y_{t-1}^2, y_{t-1}y_{t-2}, y_{t-2}^2\right]^T$. Subsequently, the parameters are estimated according to:

$$\hat{y}_i = \hat{\phi}_1 y_{i-1}^2 + \hat{\phi} y_{i-1}y_{i-2} + \hat{\phi}_3 y_{i-2}^2 + \hat{\eta}_i$$

where, $\hat{\phi}_i$ represents the model's estimated parameters, and $\hat{\eta}_i$ represents the model's estimated residuals.

We then determined the regression for the estimated residuals $\hat{\varepsilon}_i$ in $\hat{\eta}_i$ as:

$$\hat{\varepsilon}_i = \gamma_0 + \gamma_1 \hat{\eta}_{t-1} + \gamma_2 \hat{\eta}_{t-2} \cdots + \gamma_p \hat{\eta}_{t-p} + \hat{\xi}_t$$

where, γ_0 represents the estimated parameters, and $\hat{\eta}_{i-p}$ are the estimated residuals lagged in *p*. Based on the steps of Eqs. 16-18, we calculated the Tsay Test statistics, as:

$$\hat{F} = \frac{\left(\hat{\varepsilon}^{\mathrm{T}}\hat{\eta}\right)^{\mathrm{T}}\left(\hat{\eta}^{\mathrm{T}}\hat{\eta}\right)^{-1}\left(\hat{\eta}^{\mathrm{T}}\hat{\varepsilon}\right)/m}{\left(\hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon}\right)/(n-p-m-1)}$$

where, $m = p(p+1)/2$ and the null hypothesis of a linear series, *i.e.*, $H_0 : \gamma_1 = \gamma_2 = \cdots \gamma_p = 0$, is tested.

While Keenan's test and Tsay's test for nonlinearity are designed for detecting quadratic nonlinearity, they may not be sensitive to threshold nonlinearity. Here, we discuss a likelihood ratio test with the threshold model as the specific alternative. The null hypothesis is an AR(p) model versus the alternative hypothesis of a two-regime TAR model of order *p* with constant noise variance, that is; $\sigma_1 = \sigma_2 = \sigma$. With these assumptions, the general model can be rewritten as:

$$\begin{aligned} y_t &= \varphi_{1,0} + \varphi_{1,1}y_{t-1} \ldots + \varphi_{1,p}y_{t-p} \\ &+ \left\{\varphi_{2,0} + \varphi_{2,1}y_{t-1} \ldots + \varphi_{2,p}y_{t-p}\right\} I(y_{t-d} > r) + \sigma\, e_t \end{aligned}$$

where the notation $I(y_{t-d} > r)$ is an indicator variable that equals 1 if and only if the enclosed expression is true, and zero otherwise. In practice, the test is carried out with fixed *p* and *d* values. The likelihood ratio test statistic can be shown to be equivalent to:

$$T_n = (n-p)\log\left\{\frac{\hat{\sigma}^2(H0)}{\hat{\sigma}^2(H1)}\right\} \quad (17)$$

where $n - p$ is the effective sample size. The test statistic is the maximum likelihood estimator of the noise variance from the linear AR(p) fit and from the TAR fit with the threshold searched over some finite interval. Under the null hypothesis ($\varphi_{2,0} = \varphi_{2,1} = \ldots = \varphi_{2,p} = 0$) the (nuisance) parameter *r* is absent. Hence, the sampling distribution of the likelihood ratio test under H_0 is no longer approximately $\chi 2$ with *p* degrees of freedom. (18)

The results of Tests for Nonlinearity for before and after the 2008 crises (on a monthly basis) are shown on Table 1.

Table 1. Results of Tsay, Keenan and Threshold Nonlinearity test, for return of Brazilian soybean prices (on a monthly basis) for before and after 2008 crises

	Tsay´s test		Keenan´s test		Threshold	
Period	Statis-tics	*p*-value	Statis-tics	*p*-value	Statistics	*p*-value
Before 2008 crises	23.89	0.00	76.47	0.01	33.59	<0.00
After 2008 crises	0.78	0.38	0.15	0.70	46.89	<0.00

As Tsay's test for quadratic nonlinearity in a time series considers a null hypothesis that the process is linear, when we reject the null hypothesis we reject linearity for the given time series. Accordingly our results indicate the soybean market showed linearity after but not before the 2008 crisis (Table 1). Keenan's test analyses a series' non-linearity against a the null hypothesis that the time series follows some AR process. Keenan's Test shows series nonlinearity after, but not before, the 2008 crises (Table 1). In order to confirm this fact, we carried out the Threshold test for non-linearity (Table 1). The null hypothesis of the Threshold test for non-linearity is an AR(p) model *vs.* an alternative hypothesis of a two-regime TAR model of order *p* and with constant noise variance, that is; $\sigma_1 = \sigma_2 = \sigma$. This test suggests that the series returns are highly non-linear after the 2008 crisis (p <0.0001). Thus it is necessary to use a nonlinear method to forecast such a time series. Consequently a GMDH model was used.

Based on an algorithm that dates back to the 1960s, the Group Method of Data Handling (GMDH) is a mathematical method that allows one to estimate states in a system, along with controllers' exits and performers' functions (Ivakhnenko, 1969). The algorithm can be considered self-organized and of inductive propagation in the solution of practical and complex problems. Moreover, it is possible to obtain a mathematical model for a given process from data sample observations, that will be used when identifying and recognizing patterns, or even to describe the process itself.

The use of GMDH-like self-organizing networks has been successfully applied to a wide range of fields of study (Ahmadi et al., 2007). Mottaghilab et al. (2010) reported good results when this type of network was applied in specific areas, particularly such as Engineering and Economics. Most GMDH algorithms use polynomial reference functions. A general connection between entry and exit variables can be expressed by the Volterra functional series, an analogue of the Kolmogorov-Gabor polynomial:

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1} \sum_{k=1}^{n} \beta_{ijk} x_i x_j x_k + \xi.$$

where, x_i, x_j, x_k are endogenous variables, β_0, β_{ij} and β_{ijk} are the polynomial coefficients, and ξ is the error term.

The content of Ivaknenko's algorithm was developed as a vehicle to identify linear and non-linear relationships between inputs and outputs, thereby generating a structure tending towards an optimum, through a successive process of several data manipulations, via the incorporation of new layers.

The GMDH model can be analyzed as a combination of neural networks and stochastic concepts (Valença, 2005). GMDH networks are implemented with activating functions in the neurons of the hidden layers, and present a selection criterion in order to decide how many layers will be built. In the original formula, each neuron of the hidden layer to be built receives two entries and must activate a 2nd degree polynomial. As a consequence, a polynomial exit function will be generated via the combination of each pair of these entry neurons; the complexity of such a polynomial depends on the number of layers, *i.e.*, if there are two layers, we have a 4th degree polynomial function; for three layers, there will be an 8th degree function, and so on. Thus, such networks

are called polynomial, for the resulting model is a polynomial function.

For the period between October 2003 and May 2014, we carried out 127 forecasts, all for t+1 months, *i.e.*, only one step (month) ahead. It is important to note that in this period the American crisis occurred (*i.e.*, a credit crisis in the banking sector). Symptoms were however perceived in other sectors, especially the agricultural production sector.

In this regard, Krugman et al (1999) states that there is no universally-accepted formal definition for the concept of a financial crisis, but we know them when we see them. According to them, the basic element is a type of circular logic, where investors run away from an investment because they fear that it can go down, and where many, but not necessarily all pressures for the investment going down arise precisely from the flight of capital. They further note that such crises have been a recurring feature in international economy, since gold and silver coins were replaced by coin and paper.

A systemic global crisis arising in the USA strongly affected the Brazilian economy, both in terms of external trade and financial flux, particularly in terms of commercial credit lines and market application of Brazilian equity (De Freitas, 2009) . In Brazil, the most immediate effect was the downfall in stock markets, caused by significant selling off of stocks to foreign speculators that literally stepped over each other to repatriate their equity in order to cover their losses in their own countries. Consequently, there was an strong rise in the American dollar rise which directly influenced the Brazilian agribusiness sector.

In order to limit periods, we used a theoretical limit based on the work of De Freitas (2009), who stated that the period of greatest crisis-induced turbulence occurred from September 2008 to May 2009 – a period of nine months. This period was termed "during the crisis". The period consisting of the 59 months preceding the crisis (October 2003 – August 2008) was termed "before the crisis," while the "post-crisis" period was defined as occurring between June 2009 and May 2014 (59 months). Forecasts results for these periods, and overall are shown in Table 2.

Table 2: Accuracy of forecast of the logarithmic return of the soybean monthly price paid to producers in Brazil's Paraná state, during and after the 2008 American crisis,

Cat.	R^2	MSE	MAE	U	U^M	U^S	Ivakhnenko
A	0.06	0.00	0.04	0.04	0.00	0.00	0.94
B	0.01	0.00	0.05	0.05	0.00	0.00	1.00
C	0.04	0.00	0.04	0.03	0.07	0.00	0.97
D	0.17	0.00	0.03	0.03	0.01	0.00	0.85

A means full period, B means before the crisis, C means during the crisis, D means post-crisis

Based on the Ivakhnenko criterion (Eq. 8), the logarithmic return forecast for soybean price was effective for the full period, as well as during and after the crisis. Positively, we also note that the Theil U, the variance proportion (U^M) and the error bias proportion (U^S) were also adequate, indicating the absence of a systematic error in the forecast, which would denote that significant information – contained in the original series – had not been well modelled.

However, for the period prior to the 2008 American crisis, prediction results were poor, the Ivakhnenko criterion (IC = 1.0009) showing the forecasts to be unsatisfactory and the results erroneous (Table 2). Yet in the period after the crisis (June 2009 – May 2014), forecasts were satisfactory (IC = 0.8462). Based on the signals (Table 2), the forecasts were right in 71.19% of cases, and the R^2 was highest at 0.1741. Other post-crisis indicators, such as the MSE and MAE were similar or better that those for before or during the crisis.

These results denote a new behaviour of soybean prices paid to producers in Brazil after the 2007/2008 crisis. The Diebold-Mariano test shows that predictability after the 2007/2008 American crisis was greater than before the crisis (DM-statistic =2.7501 p = 0.0030). It is important to emphasise that the Diebold-Mariano test aims to test the null hypothesis of equality of expected forecast accuracy against the alternative hypothesis of one series (before

vs. after crisis) being predicted more accurately than the other. The after crisis series was shown to be predictable with the GMDH model. This research corroborates the work of Righi and Ceretta (2011), who have demonstrated that there is a mild inefficiency for the Brazilian soybean prices series, thus opening the possibility for arbitrage procedures and abnormal returns for this type of investments, as well as opportunities for the farmer to plan how to sell this commodity in moments that are more favourable.

4 Final Considerations

In the present study, we tried to assess the predictability of the monthly return of soybean price paid to producers in Brazil. The logarithmic return of this series was calculated, and the hypothesis that returns would follow a random walk, preventing predictability, was tested.

Soybean price returns show different features depending on the period – before implementing the Real plan (June 1994) and after. We used 127 months in order to simulate the modelling parameters and another 112 months to carry out forecasts (October 2003 – May 2014).

The forecast results were satisfactory for all samples. The GMDH model was able to demonstrate the returns' non-randomness, denoting inefficiency for this market, and therefore arbitrage opportunities and abnormal returns for investors, as well as the opportunity for producers in that region to plan their sales in more favourable periods.

References:

[1] Ahmadi, H, Mottaghitalab, M, Nariman-Zadeh, N (2007). Group method of data handling-type neural network prediction of broiler performance based on dietary metabolizable energy, methionine, and lysine. J. Appl. Poult. Res. 16(4): 494-501. doi: 10.3382/japr.2006-00074

[2] Ash, MS, Livezey, J, Dohlman, EN. (2006). Soybean Backgrounder. Outlook no. OCS-2006-01. USDA, Economic Research Service, Washington, D.C.. Website: http://www.ers.usda.gov/publications/ocs-oil-crops-outlook/ocs-200601.aspx (accessed 23 July 2015)

[3] Bai, J, Perron, P (1998). Estimating and testing linear models with multiple structural changes. Econometrica 66(1): 47-78. doi: 10.2307/2998540

[4] Beaver, WH (1981). Market efficiency. Accounting Review 56(1), 23-37. Jstor: http://www.jstor.org/stable/246460

[5] Black, F (1986). Noise. J. Finance 41(3), 529-543. doi: 10.1111/j.1540-6261.1986.tb04513.x

[6] Brown R, Durbin, J, Evans, JM (1975). Techniques for testing the constancy of regression relationships over time. J. Royal Stat. Soc. Ser. B (Methodological), 37(2):149-192. Jstor: http://www.jstor.org/stable/2984889

[7] Chianu, JN, Zegeye, EW, Nkonya, EM (2010). Chapter 20. Global soybean marketing and trade: a situation and outlook analysis. In: Singh, G (Editor) The Soybean: Botany, Production and Uses, Wallingford, UK, CAB International, p. 461-483. doi: 10.1079/9781845936440.0461

[8] Chow, GC (1960). Tests of equality between sets of coefficients in two linear regressions. Econometrica 28(3): 591-605. doi: 10.2307/1910133

[9] Covas, F (1997). Teste de Quebras Estruturais: uma Aplicação a Séries Portuguesas. [Testing of Structural Breaks: an Application to Portuguese Series] Universidade Nova de Lisboa, Lisbon, Portugal. (in Portuguese)

[10] Cryer, JD, Chan, K-S (2008). Time Series Analysis: with Applications in R: Springer Texts in Statistics. Springer-Verlag, New York, NY. doi: 10.1007/978-0-387-75959-3

[11] De Freitas, MCP. (2009). The effect of the global crisis in Brazil: Risk aversion and preference for liquidity in the credit market. Estudos Avançados, 23(66), 125-145. doi: 10.1590/S0103-40142009000200011

[12] Delgado, CL, Rosegrant, M, Steinfeld, H, Ehui, S, Courbois, C (1999). Livestock to 2020: The next food revolution. Food, Agriculture and the Environment Discussion Paper 28. IFPRI/FAO/ILRI, Washingtion, DC/Rome Italy/Nairobi, Kenya. Website: http://www.ifpri.org/publication/livestock-2020-0 (accessed 23 July 2015)

[13] Diebold, FX, Mariano, RS (1994). Comparing predictive accuracy: National Bureau of Economic Research Working Paper no. 169. NBER, Cambridge, MA. doi: 10.3386/t0169

[14] Dacorogna, M, Müller, U, Olsen, R, Pictet, O (2001). Defining efficiency in heterogeneous markets. Quantitative Finance 1(2): 198-201. doi:10.1080/713665666

[15] Fama, EF (1970). Efficient capital markets: A review of theory and empirical work. J. Finance 25(2): 383-417. doi: 10.1111/j.1540-6261.1970.tb00518.x

[16] FAOSTAT (2014). Website: http://faostat3.fao.org/faostat-gateway/go/to/search/soybean/E (accessed 23 July 2015)

[17] IGC (2015). Website: http://www.igc.int/en/downloads/gmrsummary/gmrsumme.pdf (accessed 26 July 2015)

[18] IPEADATA (2011). Website: http://www.ipeadata.gov.br (accessed 23 July 2015)

[19] Ivakhnenko, A (1969). Self Learning Systems for Recognition and Automatic Control. Technika, Kiev, 392 p. (In Russian).

[20] Ivakhnenko, A, Ivakhnenko, G, Müller, J (1993). Self-organization of optimum physical clustering of the data sample for weakened description and forecasting of fuzzy objects. Pattern Recognition and Image Analysis, 3(4): 415-421.

[21] Jensen, M (1978). Some anomalous evidence regarding market efficiency. J. Finan. Econ., 6(2-3): 95-101.doi: 10.2139/ssrn.244159

[22] Jun, M, Chao, S. (2010). The Study on Soybean Price Fluctuations Based on Chaos Theory. In Proc. 2nd Internat. Conf. Computer Modeling and Simulation. ICCMS '10 (January 22-24 2010, Sanya, PRC). 2: 51-55. doi: 10.1109/ICCMS.2010.266

[23] Keenan, D (1985). A Tukey nonadditivity-type test for time series nonlinearity. Biometrika, 72(1): 39-44. doi: 10.1093/biomet/72.1.39

[24] Krugman, PR, Rogoff, KS, Fischer, S, McDonough, WJ (1999). Currency Crises. In: Feldstein, M (Editor) International Capital Flows. Univ. Chicago Press, Chicao, IL, p. 421-466. Website: http://www.nber.org/chapters/c9803.pdf (accessed 23 July 2015)

[25] Malkiel, BG (2003). The efficient market hypothesis and its critics. J. Econ. Perspect. 17(1): 59-82. doi: 10.1257/089533003321164958

[26] Milionis, AE (2007). Efficient capital markets: A statistical definition and comments. Stat. Prob. Lett. 77(6), 607-613. doi:10.1016/j.spl.2006.09.007

[27] Mottaghitalab, M., Faridi, A, Darmani-Kuhi, H, France, J, Ahmadi, H (2010). Predicting caloric and feed efficiency in turkeys using the group method of data handling-type neural networks. Poult. Sci. 89(6), 1325-1331. doi: 10.3382/ps.2009-00490.

[28] Nazlioglu, S., & Soytas, U. (2010). World oil prices and agricultural commodity prices: Evidence from an emerging market. Energy Econ. 33(3): 488-496 doi: 10.1016/j.eneco.2010.11.012

[29] Pindyck, RS, Rubinfeld, DL (1991). Econometric Models and Economic Forecasts.3rd ed. McGraw-Hill, New York

[30] Righi, MB, Ceretta, PS (2011). Predictability and efficiency in agricultural market. Ciênc. Rural, 41(10): 1844-1850. doi: 10.1590/S0103-84782011005000125

[31] Rubinstein, M (1975). Securities market efficiency in an Arrow-Debreu economy. Am. Econ. Rev. 65(5), 812-824. Jstor: http://www.jstor.org/stable/1806622

[32] Singh, G (ed.) (2010). The Soybean: Botany, Production and Uses. CABI Publishing, Walllingford, UK. doi: 10.1079/9781845936440.0000

[33] Steyerberg, E (2009). Clinical Prediction Models: A Practical Approach to Development. Validation, and Updating. Springer-Verlag, New York. doi: 10.1007/978-0-387-77244-8

[34] Timmermann, A, Granger, CWJ (2004). Efficient market hypothesis and forecasting. Intern. J. Forecast. 20(1), 15-27. doi:10.1016/S0169-2070(03)00012-8

[35] Trennepohl, D, Paiva, CÁN (2011). A importância da sojicultura para o desenvolvimento da região noroeste do Rio Grande do Sul. Ensaios FEE, 31(Special no.): 741-778. (Portuguese, with English Abstract)

[36] Tsay, RS (1986). Nonlinearity tests for time series. Biometrika, 73(2), 461-466. doi: 10.2307/2336223

[37] Tsay, RS (2005). Analysis of Financial Time Series. 2nd ed. Wiley-Interscience, New York.

[38] Valença, M (2005). Aplicando redes neurais: um guia completo. Livro Rápido, Olinda–PE, Brazil.

[39] Zeileis, A, Leisch, F, Hornik, K, & Kleiber, C (2001). strucchange. An R package for testing for structural change in linear regression models. J. Stat. Software 7(2): 1-38. Website: http://www.jstatsoft.org/v07/i02 (accessed 23 July 2015).