

Autoregressive vectors model in the analysis of the determinants of soybean production in Brazil

Michael Gonçalves da Silva^I, Rafaela Goettems^{II},
Vanessa Siqueira Peres da Silva^{III}

ABSTRACT

The general objective of this work was to identify what were the determinants of soybean production increase in Brazil, between 1980 and 2015. This work uses the Self-Regressive Vector (VAR) methodology for the estimation of empirical models. Allow the obtaining of evidence capable of clarifying the question about the determinant factors of the soybean productivity in this period. They pointed evidences of a greater relation between the soybean productivity increase and the physicochemical innovations. In addition, as the data reveal, there is idle capacity to expand soybean production, as well as increased productivity for the coming years. The empirical analysis suggests that, in fact, soybean productivity in Brazil is increasing, as occurred in the forecast test performed in this study.

Keywords: Autoregressive models; Time series; Soybean Production.

^I Universidade Federal de Santa Maria, Santa Maria, Brasil. E-mail: michael_gsilva@yahoo.com.br.

^{II} Universidade Franciscana, Santa Maria, Brasil. E-mail: rafaela.goettems@hotmail.com.

^{III} Universidade Federal de Santa Maria, Santa Maria, Brasil. E-mail: dexvanessa@gmail.com.



1 INTRODUCTION

Brazil is one of the largest soybean producers in the world and is the country that has the greatest potential for expansion in cultivated area, and may become the largest producer and exporter of soybean and its derivatives in the coming years Viera e FISHLOW (2017).

Soybean farming in Brazilian lands began in the last century, at which time soybeans were produced mostly for domestic consumption and on small rural properties. The change of landscape occurred in the 1950s, when the Federal Government encouraged the planting of wheat in the country, generating the need to find a legume to perform the crop rotation regime and in the early 1960s, soy began to exercise this function. The wheat-soybean binomial presented high viability, as there was the possibility of sharing soil, inputs and machines. The considerable increase in world demand for food has made room for Brazil to compete in the commodity export market. In order to play a relevant role in global supply, however, Brazilian agriculture had to seek evolutions with the use of technologies and the expansion of productive frontiers. This allowed the soy to gain scale, leading the country to reach the level of agricultural power Vieira e FISHLOW (2017).

Since then, the soy complex has played an important role in the development of the Brazilian economy. The growth of soy-related sectors through investments in technology, new agricultural areas, and grain processing and oil refining industries has promoted positive results not only in operating volumes but also in improving the population's life.

According to Silva (1990), a country has two basic alternatives to increase its agricultural productivity: i) intensify production in already used lands, through the incorporation of technological progress; ii) expand the areas used by incorporating new land into the agricultural frontier.

Thus, the aim of this study is to bring information about Brazilian agriculture, especially about soybean, more specifically in soybean, investigating the evolution of the agricultural processes, the progress in the use of technologies, as well as the increase of the Brazilian agricultural frontier. Based on these initial conjectures, this

paper raises the following problem: What are the determinants of the increase in Brazilian soy productivity since the 1980s?

This research is justified by the economic importance of Brazilian soybean, both internally and externally. After its introduction in Brazil, still in the nineteenth century, it remains ignored for about 70 years, being used only as cattle fodder or as animal fattening in small properties. However, from the 1960s onwards, an unprecedented growth trajectory began in the country's history, becoming, in less than twenty years, the leading culture of Brazilian agribusiness. Since then, year after year, record soybean production marks have been registered, which has made Brazil the second largest oilseed producer in the world, with real chances of becoming the largest in a few years. Thus, it is essential to find out which elements involved in production have been responsible for the exponential increase in production in recent years.

2 METHODS

This paper uses the Self-Regressive Vector (VAR) methodology to estimate empirical models that allow the obtaining of evidence capable of clarifying the question about the determinant factors of soybean productivity increase between 1980 and 2015.

2.1 Research Classification

Scientific research can be classified under several approaches. Silva (2014) clarifies that the classification varies according to the objectives that were defined by the researcher.

This research is characterized by its applied purpose is aimed at the acquisition of knowledge for application in a specific situation. The present work configures a descriptive objective research with quantitative approach. Descriptive because it aims to describe, analyze and verify the relationship between facts and also the establishment of relationship between the variables, and quantitative because it is characterized by the use of quantification, both in the modalities of information collection and in their treatment through statistical techniques SILVA (2014).

There is also the classification of the methods, which are responsible for identifying the reasoning used in the development of the research. The work developed here uses the inductive method. The induction starts from a phenomenon to reach a general law through observation and experimentation of the data, aiming to find out if there is relationship between the two phenomena.

It is also characterized by using bibliographic and documentary technical procedure. Regarding data collection, Silva (2014) points out that data sources can be primary or secondary. The data classified as secondary are those that have already been collected for purposes other than those of the problem at hand. For this reason, data from this study are considered secondary.

2.2 Methodology, Data Source and Description of Estimated Models

The autoregressive vectors econometric model - VAR, using free software GRETL (2019) (Gnu Regression, Econometrics and Time-series Library) for calculation resolution and statistical analysis.

The VAR methodology is commonly used to construct interrelated time series forecasting systems and to analyze the dynamic impacts of random disturbances on the system of variables that make up the model. For this, the VAR procedure treats all variables as endogenous, where the value of each variable is expressed as a linear function of the lagged values of itself and all other variables included in the model Bueno (2011).

According to Bueno (2011), the model is an interesting and effective way to characterize the dynamic interactions between economic variables, without relying on theoretical constraints that a traditional econometric model would have. The VAR methodology becomes useful to this study, as it is possible the interaction of some explanatory variables of soybean yield, without previously assuming a causal relationship between the variables that compose the model.

The overall VAR model has many parameters, and can be difficult to interpret because of the complex interaction and feedback between variables in the model. According to Bueno (2011), the dynamic properties of a VAR are often summarized

using various types of structural analysis, the main types being variance decomposition (VDC) and observing response functions. Impulse Response Function (IRF).

Bueno (2011) expresses the generalization of an autoregressive model of order p by a vector with n endogenous variables, X_t , being connected to each other by a matrix A , expressed below:

$$AX_t = B_0 + \sum_{i=1}^p B_i X_{t-i} + B\varepsilon_t \quad (1)$$

The estimation of the VAR models proposed in this work involves the use of annual data for the period from 1980 to 2015, obtained from the statistics of the National Supply Company – CONAB (2019), National Association of Motor Vehicle Manufacturers – ANFAVEA (2019), National Association for Diffusion of Fertilizers. – ANDA (2019) and Céleres (2019) company. The difficulty of obtaining accurate information before and after the specified years justify to study the period. However, considered this period is adequate for the estimation of the considered models, considering that it encompasses the period of agricultural industrialization, the expansion of the productive frontier to the Brazilian Midwest, as well as the release of production with transgenic seeds.

The variables used in the estimates were:

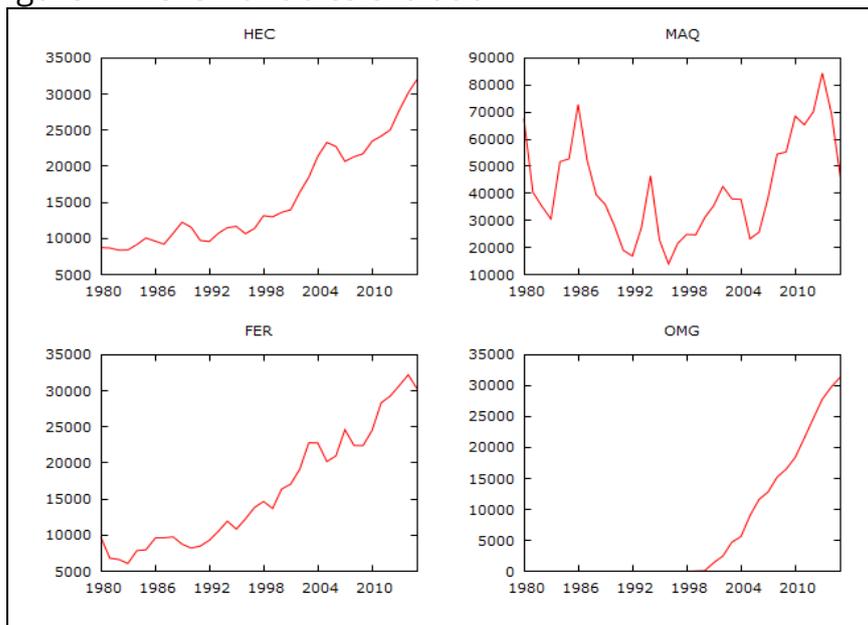
- PTS = Soybean crop productivity
- MAQ = Agricultural machines marketed as a proxy for mechanical innovations.
- FER = Fertilizers delivered to the consumer as a proxy for physicochemical innovations.
- OMG = Hectares planted with genetically modified organisms as a proxy for biological innovations.
- HEC = Planted area as a proxy for agronomic innovations.

3 DISCUSSION AND RESULTS

3.1 Stationary Test and Selection of VAR Model

The estimation of the VAR methodology assumes the performance of series stationarity tests. Enders (2010) notes that the stationarity of a time series can be visually verified by its correlogram and formally tested unit root tests. Figure 1 shows the trends of the level variables listed during the analyzed period.

Figure 1 - Level variables evolution

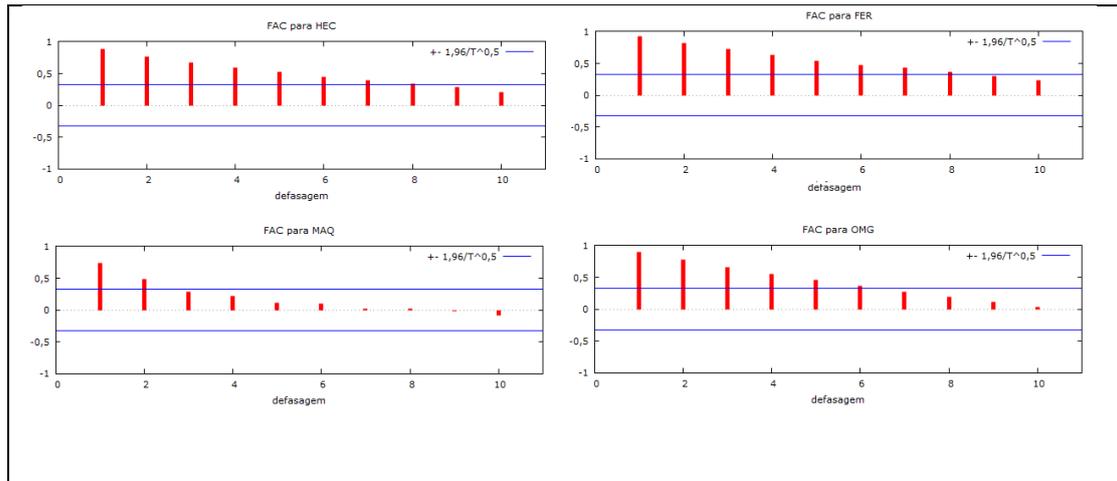


The series stationarity diagnosis used in this study was based on the autocorrelation function test (FAC) and the resulting correlogram, which is simply the graphical representation of the FAC against the size of the lag. The sample autocorrelation function in the lag is:

$$\hat{\rho}_k = \frac{\hat{y}_k}{\hat{y}_0} \quad (2)$$

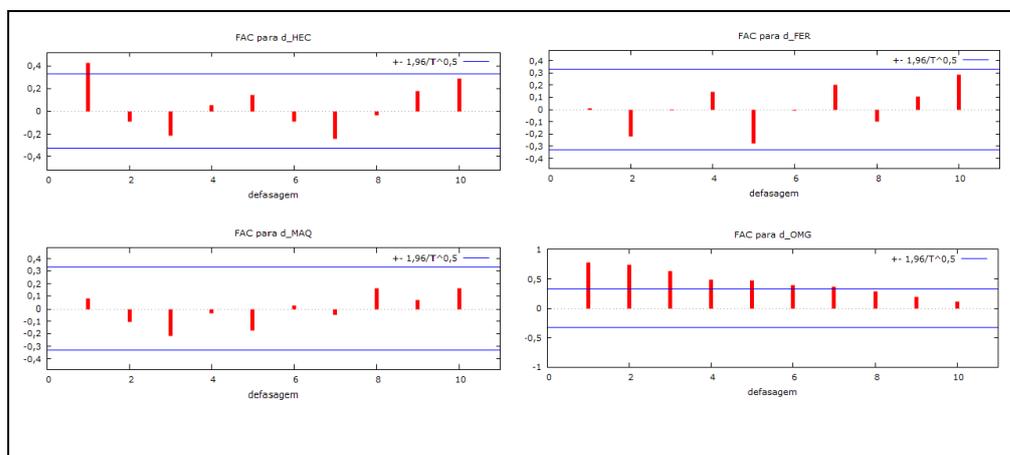
It is simply the ratio between the sample covariance and the sample variance. How much more approaches 1, the greater the indication that the time series is non-stationary. Another indicator of nonstationarity is the graphical representation of k , known as the correlogram. When the correlogram starts at a very high value (close to 1) and gradually decays, there is a strong indication of a nonstationary time series (GUJARATI, 2000). The graphical representations of the FAC are expressed in Figure 2.

Figure 2 – Graphical Representation of FAC (correlogram)



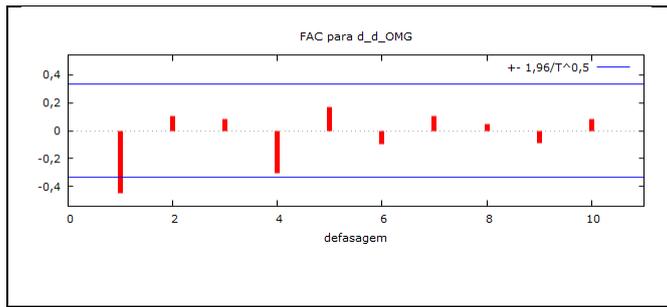
As a result, it can be inferred that this series has seasonal unitary root. Based on the results of the unit root tests with the level variable, to eliminate this seasonal behavior, an order 1 difference was applied. Next, Figure 3 shows the correlate of the autocorrelation function, but this time with the application of an order difference 1.

Figure 3 – Graphical Representation of FAC in First Difference



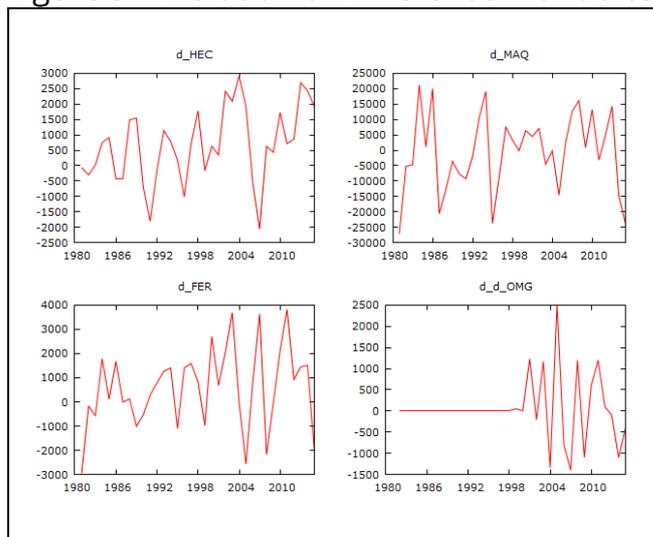
After performing the autocorrelation function test and the analysis of the resulting correlogram, it is assumed that the HEC, MAQ, and RES variables of the model should be taken in the first difference, however, the OMG variable was better adjusted when in the second difference eliminate seasonality from the model, as shown in Figure 4. Figure 5 below shows the trends of the variables in differences over the period analyzed.

Figure 4 - FAC Graphical Representation for Second Difference OMG



From the results of the unit root tests, we proceeded to the specification and choice of the model lags, whose results are shown in Table 1. The model was estimated using four lags in order to verify the most appropriate lag to explain the model in question. The choice of the appropriate model was made based on the best values presented by the Akaike criterion (AIC), Schwarz Bayesian criterion (BIC) and Hannan-Quinn criterion (HQC).

Figure 5 - Evolution of Differential Variables



Proceeding in this way, it was chosen as the best option for the study, based on the principle of parsimony, the use of the VAR system with only one lag, by the principle of parsimony, represented by VAR (1).

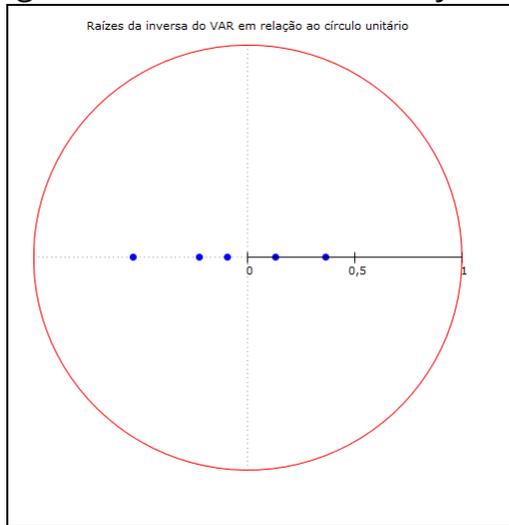
Table 1- Choice of VAR lag

Phasing	AIC	BIC	HQC
1	86,161555	87,562752*	86,609810
2	86,061416	88,630278	86,883217
3	86,087349	89,823876	87,282696
4	83,576115*	88,480306	85,145008*

*valor nominal de significância de 5% ($P < 0, 05$).

The diagnostic tests were performed to verify the stability of the VAR models and this stability check is possible through the inverse roots of the autoregressive polynomials, which must be within the unit circle, as shown in Figure 6.

Figure 6 - Inverse Roots of Polynomials



It is observed that all inverse roots of the model polynomials were inside the circle, indicating that the values are smaller than the unit. Thus, it adds greater robustness to the estimation performed.

3.2 Forecast Analysis Results, Impulse Response Functions, and Analysis of Variance Decomposition

According to the purpose of this work, the estimation of the impulse response functions and the error variance decomposition analysis involve the evaluation of a possible evidence about what determined the soybean productivity increase in Brazil. In addition to those already mentioned, the forecast analysis of the VAR model will be performed.

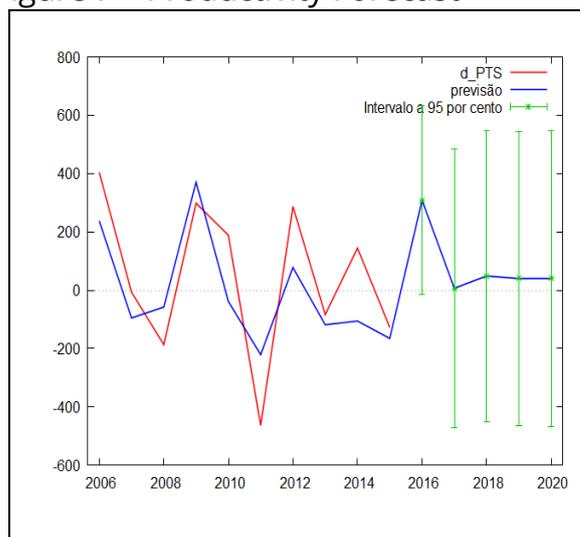
It is worth mentioning here the limitations of the proposed model, since it is difficult to analyze agricultural productivity only by technical and genetic bias, without taking into account climatic factors (temperature, radiation and water requirement), production techniques (grain variety, plant population, sowing date) and edaphic - nutrients and soil pH, among others.

3.2.1 Model Forecast Analysis

One of the main purposes of multivariate time series analysis is to generate prediction. The methodology called Autoregressive Vectors (VAR), allows the analysis of possible relationships between the series and the dynamic relationships that occur between them, thus increasing the quality of the model and the predictions. According to Morettin and Tolo (2004) one can make predictions of future values of the series, being these short or long term predictions.

Howevwe the Brazilian soybean yield (PTS) was forecast for the next years, which is shown in Figure 7. These forecasts were made within the observed period, reserving the last ten observations (crop years 2006 to 2015) so that it was possible to draw comparisons between the values and to be able to evaluate the model used. In addition, the same variable was predicted by 2020.

Figure 7 – Productivity Forecast



The prediction results indicate that the proposed model is capable of producing satisfactory predictions within the proposed one, in order to determine the soybean crop yield projection, which is corroborated by the overestimated variation analysis presented in Table 2. This model is corroborated by the fact that a multivariate model was used. Thus, the model used proves to be an efficient forecasting technique, as the predicted values are very close to the actual values.

Table 2 – Overestimated Productivity Change

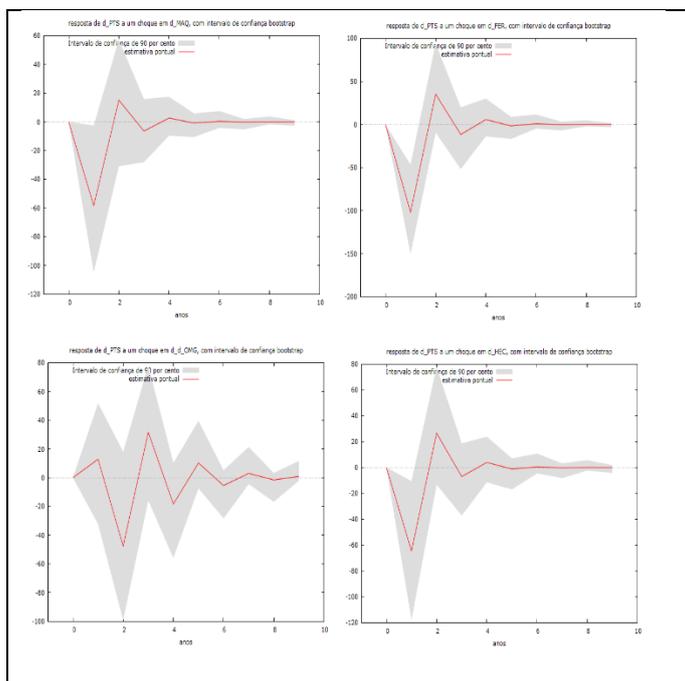
Ano	PTS observado	PTS projetado	Variação superestimada
2010	3115,00	2792,26	-10,36%
2011	2651,00	2834,76	6,93%
2012	2938,00	2778,51	-5,43%
2013	2854,00	2934,62	2,82%
2014	2998,00	2852,66	-4,85%
2015	2870,00	3111,79	8,42%
2016	3364,00	3252,05	-3,33%
2017	3394,00	3403,90	0,3%
2018	3302,00	3466,35	4,98%
2019	nd	3498,67	nd
2020	nd	3536,35	nd

3.2.2 Impulse Response Function Analysis

The analysis of impulse response functions (FIR) aims to determine the direction of the effects of shocks in time “t” of the error terms of a particular variable on the values of the other variables that make up the endogenous structure of the VAR.

Corroborating the objective of this research and determining factors for the increase of Brazilian soybean productivity, the figure below presents the results of the FIR related to the productivity responses in relation to mechanical innovations (Effect I), physicochemical (Effect II) and biological (Effect III) renewals and the restructuring of agronomic processes (Effect IV). It is noteworthy that a horizon of ten periods was used for the diagnosis. Figure 8 indicates the graphical results of the tests.

Figure 8 – Impulse Function Answer on Variables



According to the impulse response function, when analyzing a shock of traded agricultural machinery (d_MAQ) on total soybean yield (d_PTS), a negative response can be verified in the first period and then reversed in the second period. Analyzing the response and a shock of traded fertilizers (d_FER) on total soybean yield (d_PTS), one finds again a negative response in the first period and a positive response in the second period.

Considering the shocks of the adoption of GMOs (d_d_OGM), it is possible to see a positive response in the first period, then in period 2 it takes a negative and again positive trend in period 3.

Finally, with a planted area shock (d_HEC) it gets negative response in the first period and positive in the second, with slight oscillation in periods 3 and 4.

3.2.3 Analysis of Variance Decomposition

Analysis of Variance Decomposition (ADV) is another instrument used to describe system dynamics under the Autoregressive Vector model. According to Bueno (2011) it is a way of identifying the responsibility of each variable in explaining the variance of all system variables, serving as a classification of the relative importance of each variable in determining itself and the others.

The results of the variance decomposition of the PTS prediction errors showed that, after 13 periods after an unanticipated shock on this variable, approximately 57% of its behavior is due to itself, and approximately 43% is attributed to the other variables, as shown in the Table 3.

Table 3 – Decomposition of variance

Período	Erro	d_PTS	d_HEC	d_MAQ	d_FER	d_OMG
1	157,401	100,0000	0,0000	0,0000	0,0000	0,0000
2	232,748	62,3901	7,3127	10,6404	19,3523	0,3045
3	243,13	58,8930	7,3909	10,2445	19,3549	4,1167
4	246,234	57,7481	7,3126	10,0814	19,1966	5,6613
5	247,196	57,3799	7,2916	10,0234	19,1373	6,1677
6	247,472	57,2718	7,2823	10,0055	19,1151	6,3253
7	247,552	57,2403	7,2799	10,0003	19,1090	6,3705
8	247,575	57,2313	7,2792	9,9988	19,1071	6,3837
9	247,581	57,2287	7,2790	9,9983	19,1066	6,3874
10	247,583	57,2280	7,2789	9,9982	19,1064	6,3885
11	247,584	57,2278	7,2789	9,9982	19,1064	6,3888
12	247,584	57,2277	7,2789	9,9982	19,1063	6,3889
13	247,584	57,2277	7,2789	9,9982	19,1063	6,3889

Source: Own elaboration from GRETL results.

According to the ADV results, the combination of increased soybean yield in the period analyzed is 19,1% physicochemical innovations; 9,99% mechanical innovations, 7,28% agronomic innovations and 6,39% for biological innovations.

In contrast to what Silva (1990) states by saying that “biological and agronomic innovations form the basis of technical progress for agricultural development”, the econometric test presented above shows that physicochemical innovations are more closely related to the increase of soybean yield in the analyzed period.

4 CONCLUSIONS

The importance of agriculture for the Brazilian economy has been widely discussed in the economic literature, considering that this is one of the largest employment and income generating sectors in the country, besides having an active role in the Brazilian export agenda Vieira e FISHLOW (2017).

In the last decade in Brazil, productivity has been gaining importance in determining economic growth. Productivity is a central theme in the debate on economic development, because it is the factor that determines output growth in the long run. Regarding agriculture, performance in recent years has been a success. However, the increase in agricultural productivity demanded knowledge, which in turn, education and research. These last ones bring innovations. Over the last few decades, soybeans have spread throughout Brazil's fields. Secondary farming in the 60s, became one of the highlights of the country's agribusiness. Crop was introduced to Brazilian

agriculture still as forage in the late nineteenth century and after more than fifty years fulfilling this role, with the emergence of cattle raising on an industrial scale, soybean led the largest expansion ever occurred with any crop. Since 1970, more than tripled the area of land devoted to soy cultivation.

Brazil occupies for some years the position of second largest producer of soybean, presented each year breaking record production. However, the increase in production was not only due to the increase in area, but also due to the increase in productivity, which is a measure of efficiency, because it represents more kilograms of grain per unit area. In this context, Silva (1990) discusses four factors that influence the improvement in crop productivity, namely: i) agronomic innovations, which allow new methods of production organization, increasing labor productivity in general; ii) biological innovations, which reduce the production period and / or enhance the effects of mechanical and physicochemical innovations; iii) physicochemical innovations, which modify the natural conditions of the soil; iv) mechanical innovations, which reduce the time and labor employed in a given work.

Thus, this research aims to contribute to a better understanding of the factors that determined the improvement in soybean yield in Brazil, by studying the potential impact of the innovations discussed by Silva (1990). Thus, we sought to develop a model based on the methodology of Auto-Regressive Vectors (VAR), Impulse Response Function (FIR) and Analysis of Variance Decomposition (ADV), which allowed investigating the relationships between the productivity of Brazilian soybean. and agronomic, biological, physicochemical and mechanical innovations in the 1980-2015 period.

Therefore, an analysis was made of each of the proposed innovations, verifying that all of them went through a great process of beneficitation, diffusion and innovation. Moreover, since the variable of interest is soybean yield, the effect of shocks on soybean was observed considering the simultaneous effects of all variables included in the system. Graphical analysis suggests that the introduction of biological innovations (transgenic seeds) would have a positive impact, while mechanical, agronomic and physicochemical innovations (machinery, new areas and fertilizers respectively) would

negatively affect soybean yield at the time of shock. However, in the analysis of the significance of the influence of shocks from one of the variables of interest on the variance composition of another variable, it is possible to observe that the decomposition of variance shows the soybean yield in Brazil is strongly independent of the other variables since productivity is explained by itself at a level of 57%.

When analyzing the 43% that are not explained by the productivity itself, the ADV econometric tests showed evidence of a greater relationship between the increase in soybean productivity and the physicochemical innovations for the analyzed period, since it was found that variable explains in 19,1% the improvement in productivity.

In addition, as the data reveal, there is idle capacity to expand soybean production, and there is a forecast of increased productivity for the coming years. The empirical analysis suggests that, in fact, soybean productivity in Brazil is increasing, as occurred in the forecast test performed in this study.

Observing the aspects presented, that soybean cultivation is extremely important for both national and international agriculture, in view of its great versatility, being marketed in nature, bran, farina, oil, fuel, feed and bark among others. Both production and consumption of soy-related products are expected to grow further in the coming years, based on population growth, increased use of soy as raw materials for industry, increasing consumption for the meat industry. The challenge is clear: we will plant more soy and we will need more land to grow. Thus it is necessary to find methods and ways to continue producing more with less resources, in view of the sustainability of the ecosystem and the planet.

The model presented plays an important role in determining the variables that increase soybean yield in Brazil. Thus, the application of this model extends to other agricultural commodities, thus contributing to the investigation of the variables responsible for the increase in productivity in other crops, which in turn may contribute to the reduction of production costs by increasing yield gains. scale.

REFERENCES

ANDA. Associação Nacional para Difusão de Adubos. Estatísticas. Planilhas. Disponível em: <http://www.anda.org.br/index.php?mpg=03.01.00&ver=por>. Acesso em: 30 mai. 2018.

ANFAVEA – Associação Nacional dos Fabricantes de Veículos Automotores. **Anuário da Indústria Automobilística Brasileira 2007**. Acesso em: 21 abr. 2018.

ANFAVEA - Associação Nacional dos Fabricantes de Veículos Automotores. Estatísticas. Disponível em: <http://www.anfavea.com.br/estatisticas.html>. Acesso em: 20 abr. 2018.

BUENO, R.L.S. **Econometria de Séries Temporais**. São Paulo: Cengage Learning, 2011.

CÉLERES. **Informativa Biotecnologia** 2017. Disponível em: <http://www.celeres.com.br/3o-levantamento-de-adocao-da-biotecnologia-agricola-no-brasil-safra-201617/>. Acesso em: 27 mai. 2018.

CONAB. Conab – Companhia Nacional de Abastecimento. **Series históricas**. Disponível em: <http://www.conab.gov.br/conteudos>. Acesso em: 25 mar. 2018

GUJARATI, D.N. **Econometria básica**. 3. ed. São Paulo: Makron Books, 2000

MALHOTRA, N. **Pesquisa de marketing**. 3. ed. Porto Alegre: Bookman, 2001.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. 1. ed. São Paulo: Editora Edgard Blucher Ltda., 535 p., 2004.

RICHARDSON, R. J. Pesquisa social: métodos e técnicas. 3. ed. São Paulo: Atlas, 1999.

SILVA, J. G. **Progresso técnico e relações de trabalho na agricultura**. São Paulo: HUCITEC, 1990.

SILVA, A. J. H. **Metodologia de pesquisa: conceitos gerais**. Guarapauva: Editora Unicentro, 2014.

VIEIRA J. E. R, FISHLOW, A. **Agricultura e indústria no Brasil: inovação e competitividade**. Brasília: IPEA, 2017.