



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

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

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

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

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

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

IMPACT of the Expected CROP YIELD On the Soybean MARKET In Parana

**Andréia Cristina de Oliveira Adami, Center for Advanced Studies on Applied
Economics - Cepea, Brazil, deiaadami@terra.com.br and Vitor Augusto Ozaki,
University of Sao Paulo, Piracicaba, Brazil, vitorozaki@gmail.com**

**Selected Poster prepared for presentation at the International Association of
Agricultural Economists (IAAE) Triennial Conference, Foz do Iguaçu, Brazil, 18-
24 August, 2012.**

**Copyright 2012 by Andréia C. O. Adami and Vitor A. Ozaki. All rights reserved.
Readers may make verbatim copies of this document for non-commercial purposes
by any means, provided that this copyright notice appears on all such copies.**

IMPACT OF THE EXPECTED CROP YIELD ON THE SOYBEAN MARKET IN PARANA

ABSTRACT

The objective of this study was to assess the impact of variations in the area and yield of soybeans planted in the state of Parana on the price in spot market. Vector Auto Regression-VAR methodology was used to set a dynamic model in order to evaluate the interaction between the three chosen variables and study expected effects over time. In cases of severe losses, the agents will only be able greater certainty about the information after dissemination of data from the IBGE and thus the impact on prices the first time will be partial, complete transmission of the impact on the physical market will only occur after the disclosure of monitoring performance and area by the IBGE. To determine the temporal relationships of precedence among the three variables we applied the Granger causality test. As a result of this test, two significant relationships at the 5% significance level with two lags were obtained. Thus, in terms of temporal precedence, variable expected yield is important to help explain the behavior of the spot market price of soybeans in Parana, the other relationship - price on area - are also important to help explaining the decision to increase or not the soybean area based on the price. The equation for growth rate of the VAR model indicate that only de expectation yield and price (5% significance), both in the first time difference, were important to predict the prices behavior on the physical market. About yield, no one variable were significant in explain their behavior. As for area with soybeans in Parana, the constant model and the first time price difference are important to help explain the growth of the soybean area in the state, also at 5%.

Keywords: soybeans, expectations, physical market

INTRODUCTION

The monitoring of agricultural yield is of fundamental importance because it provides information on the conditions of vegetative growth of the plant to help identify losses that may occur during the development cycle because of adverse weather events (droughts or excessive rain, for example).

Adverse weather events can greatly reduce the agricultural yield (supply shocks). This reduction in crop yield directly impacts the price in the spot market, increasing market risks and may cause serious financial damage to economic agents.

Soybean is the main grain produced in Brazil with an estimated production for the 2010/2011 harvest of 75,324,300 tons. Approximately, 20% of production is concentrated in the state of Parana. Thus variations in the crop yield expectations in this state impact prices in the spot market. In this scenario the objective is to evaluate the impact of changes in the planted area and the expectation of the soybean yield on the spot price in the state of Parana.

The question is how prices can be affected by the expectations of the planted area and yields in order to anticipate possible losses that may occur or anticipate possible adverse reactions of the agents in the face of production information.

We model the data using a dynamic model of time series (Vector Auto Regression-VAR) to evaluate the interaction between these variables and study the expected effects over time. Our hypothesis is based on the fact that positive changes in the expected yield and acreage have the effect of causing negative changes on spot

prices. However, there is a delay in releasing the agricultural data. The Brazilian Institute of Geography and Statistics (IBGE) is responsible for these statistics.

The economic agents can only update their expectations with a lag in time. Thus the impact on prices will be with some months of delay, for example, the market can only revise their expectations after the IBGE release the data, which happen one month later. In this context, the price in the current month should reflect the conditions of the market in the current month and the expectations of area and yield disclosed in the current month, but related to the previous month in which the data were collected.

METHODOLOGY

Several studies in the economic literature have been developed to analyze causal relationships in agriculture. Many of these studies were based on tests of causality proposed by Granger (1969). According to the construction of the Granger causality test, a causal relationship can be found if and only if, past values of one variable (X) help predict the values of another variable (Y).

The causality test in this case is based on the joint significance of the coefficients associated with lagged (past) explanatory variable. Mafioletti (2000) studied the relationship between the monthly price of soybeans, meal and oil at different levels in the brazilian market (producer, wholesaler and consumer) and between domestic and international prices. Moraes (2002) examined the causality between soybean prices in the brazilian and international market prices.

Studies using time series models to analyze the transmission of prices between agricultural markets are common in national and international literature. Alves (2002) analyzed the price transmission between the markets of the main products of this sector in São Paulo using VAR (Vector Auto Regression) models with vector error corrections

(VEC). In this study construction of VAR model and Granger causality test were the tools used to find causal relationships between variables and the impact on prices in the physical market.

ECONOMETRIC PROCEDURES

Cointegration and error correction model

The procedure used to evaluate long-term relationships between a set of variables is based on cointegration tests. To proceed to the cointegration test is necessary if variables are nonstationary and integrated of the same order. Thus, it is necessary that unit root tests be performed in the series to set the order of integration between variables.

As defined by Engle and Granger (1987), a series with no deterministic component, with an ARMA (Autoregressive and Moving Average) representation, stationary and invertible, after d differences, is said to be integrated of order d , denoted by $x_t \sim I(d)$.

Thus order of integration refers to the number of times a temporal series need to be differentiated and to become stationary (Fuller, 1976; Dickey and Fuller, 1979; Dickey and Fuller, 1981). It is known that if the generating process of a given time series is stationary, its characteristics do not change over time. In other words, if its mean and variance are constant over time and the covariance between the values of the series depends only on the lag (t).

The hypothesis test is based on the distributions of Dickey & Fuller (1979), Dickey & Fuller (1981), Fuller (1976). To check the stationarity of a given time series, we used the procedure proposed by Enders (2004). The first unit root test was

developed by Fuller (1976), considering an autoregressive process of order one [AR (1)], as described below - equation (1):

$$x_t = \rho x_{t-1} + \varepsilon_t \quad (1)$$

In equation (1) ε_t is considered white noise. The null hypothesis is that x_t is not stationary. Thus, we have that: $H_0: \rho = 1$ against $H_A: \rho < 1$. This is equivalent to testing: $\Delta y_t = (\rho - 1)x_{t-1} + \varepsilon_t$ the case $H_0: \rho = 1$ against $H_A: \rho < 1$.

The non-rejection of the null hypothesis indicates that the process has a unit root and therefore is not stationary.

Considering models incorporating the presence of an intercept and trend - equations (2) and (3):

$$x_t = \alpha + \rho x_{t-1} + \varepsilon_t \quad (2)$$

And

$$x_t = \alpha + \beta t + \rho x_{t-1} + \varepsilon_t \quad (3)$$

The statistics used in the presence of an intercept is τ_μ , and to test for trend, we use the τ_τ statistics. It is also possible to test altogether and these tests are known as ϕ and correspond to an F test. In the case of the test called ϕ_1 , it tests the hypothesis that $(\alpha, \rho) = (0, 1)$ against the hypothesis that $(\alpha, \rho) \neq (0, 1)$. In the case of ϕ_2 the null hypothesis is that $(\alpha, \beta, \rho) = (0, 0, 1)$, against the alternative hypothesis that $(\alpha, \beta, \rho) \neq (0, 0, 1)$. Finally, the ϕ_3 statistic tests the null hypothesis $(\alpha, \beta, \rho) = (\alpha, 0, 1)$ against the alternative hypothesis that $(\alpha, \beta, \rho) \neq (\alpha, 0, 1)$. The critical values for these tests are tabulated in Dickey and Fuller (1981).

You must set the order of the autoregressive process p (number of lags statistically significant) that describes the behavior of the series - equation (4).

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \sum_{i=1}^{p-1} \lambda_i \Delta x_{t-i} + \varepsilon_t \quad (4)$$

Where, $\lambda_i = \sum_{j=i+1}^p \rho_j$ and $\gamma = \sum_{i=1}^p p_i - 1$. In this case, the presence of unit root is tested by the $H_0: \gamma = 0$ hypothesis. This test is called Augmented Dickey-Fuller Test (ADF). To help identify the stationarity of the series and autoregressive terms (number of lags) the auto correlation and partial auto-correlation can be used. To determine p - order of the autoregressive process - some criteria such as Akaike (Akaike Information Criterion - AIC) and Schwarz (Schwarz Bayesian Criterion - SBC) can also be used. The AIC and SBC criteria are defined as - equations (6) and (7):

$$AIC = T \ln(sqr) + 2n \quad (6)$$

$$SBC = T \ln(sqr) + n \ln(T) \quad (7)$$

In addition to these two criteria Q statistic Ljung and Box (1978) is also used to check for serial correlation. The statistic Q is represented by equation (8):

$$Q = T(T+2) \sum_{k=1}^s r_k^2 / (T-k) \quad (8)$$

In equations (6), (7) and (8) n is the number of the estimated parameters, T is the number of observations, sqr is the sum of squared residuals of the model with lag p , r_k is the auto-correlation for lag k and s is the number of lags to be tested. The importance of cointegration analysis arises from its use in non-stationary series, thus removing the tendency (by differentiation) between of long-term variables. If two variables have a long-run equilibrium relationship, even if they have stochastic trends, they will move together over time and their difference is stable. In the cases of VAR type models estimations containing nonstationary variables, there may be stationary

linear combinations for integrated variables of the same order, for example, long-term equilibrium relationships that must be included in the model to avoid errors of estimation. Therefore, one can use the long-term stability of the comoviments between series for modeling and forecasting. These comoviments (cointegration) generate an Error Correction Mechanism (ECM) of short-term random deviations that must be included in the model. The new model to be estimated is a VAR model with error correction or VEC.

Consider two nonstationary series X_{1T} and X_{2T} , ie, stochastic processes with independent innovations. Therefore, any attempt to predict their future values using past values of the series will be meaningless - equations (9) and (10):

$$x_{1t} = x_{1t-1} + e_{1t} \quad (9)$$

$$x_{2t} = x_{2t-1} + e_{2t} \quad (10)$$

However, if there is an stable long-term relationship between X_{1T} and X_{2T} (the series are integrated of the same order and cointegrated) the difference between them is stable. This difference, which can be represented by: $z_t = X_{1T} - X_{2T}$ defines a stationary linear combination between these two variables. In matrix notation: $z_t = \beta' X_t$.

In this case, β is known as a vector error correction or cointegration vector and z_t is the mechanism for error correction or error correction model, which describes the dynamics of convergence of series in the long run.

Formally, Engle and Granger (1987) considered a set of economic variables (system) in long-run equilibrium when - equation (11):

$$\beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_n x_{nt} = 0 \quad (11)$$

In equation (11) β and X_t represent the vectors $(\beta_1, \beta_2, \dots, \beta_n)$ and $(x_{1t}, x_{2t}, \dots, x_{nt})$. The system is in long-term equilibrium when $\beta' X_t = 0$. Deviations from the long-run equilibrium errors are called equilibrium errors and are represented as - equation 12:

$$e_t = \beta' X_t \quad (12)$$

Once they are deviations from long-term equilibrium and therefore temporary, e_t is stationary. The components of the vector X_t are cointegrated of order b, d , or $x_t \sim CI(d, b)$ if all the vector components are integrated of order d , if there is a vector so that there is an integrated linear combination of order $(d-b)$ where $b > 0$, which means that the resulting linear combination (z_t) has order of integration smaller than the original variables. In this case, the vector is called the cointegrating vector.

In the short run cointegrated variables suffer deviations from long-term relationship. Without specifying a dynamic model, it is not possible to determine how the adjustment will occur. This problem would be solved by applying an error correction model, so that the deviation is corrected in the previous period (Enders, 2004).

In a system composed by more than two integrated series of the same order, cointegration can be tested using the method proposed by Johansen (1988). This method is a multivariate version of the method of Engle and Granger for the detection of cointegration for two variables and it uses maximum likelihood estimation to test for and estimate cointegration vectors.

This procedure focuses on the relationship between the rank of a matrix and its characteristic roots. Consider the case of n -variable - Equation (13):

$$X_t = A_1 X_{t-1} + \varepsilon_t$$

$$\Delta X_t = (A_1 - I) X_{t-1} + \varepsilon_t$$

$$\Delta X_t = \pi X_{t-1} + \varepsilon_t \quad (13)$$

In equations (13), X_{t-1} and ε_t are vectors ($n \times 1$); A_1 is the matrix of parameters ($n \times n$); and π is defined as $(A_1 - I)$ and I an identity matrix ($n \times n$). The rank is equal to the number of cointegrating vectors. Assuming that there are no linear combinations that are stationary and therefore the variables are not cointegrated. To know the number of cointegrating vectors, one must check the significance of the characteristic roots π . The test to determine the number of characteristic roots that are not significantly different from zero is accomplished through two statistics - equations (14) and (15):

$$\lambda_{\text{traço}}(r) = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i) \quad (14)$$

$$\lambda_{\text{max}}(r, r+1) = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_{r+1}) \quad (15)$$

In equations (14) and (15) $\hat{\lambda}$ are the estimated values of the characteristic roots obtained by the estimation of the matrix π and T is the number of observations. To verify the number of lags required for the model multiecuacional one can use the *AIC* criterion which consists of - equation (16):

$$\text{minimize } AIC = \left| \tilde{\Omega} \right| \exp[(2u)/T] \quad (16)$$

Granger Causality

For two time series X_t and Y_t , the Granger causality test assumes that the relevant information for the prediction of the respective variables X and Y is contained only in the time series on these two variables. Therefore, a stationary time series X causes another series Y if better predictions statistically significant of Y can be obtained by including lagged values of X to the lagged values of Y. Thus, the test involves estimating the following regressions given by equations (19) and (20):

$$X_t = \sum a_i Y_{t-i} + \sum b_i X_{t-i} + u_{1t} \quad (17)$$

$$Y_t = \sum c_i Y_{t-i} + \sum d_i X_{t-i} + u_{2t} \quad (18)$$

In equations (17) and (18) u_{it} are the noise, which is assumed uncorrelated. One can distinguish four different cases of Granger causality:

Unilateral causality from Y to X, when the estimated coefficients in (17) for the lagged variable Y are jointly different from zero ($\sum a_i \neq 0$) and when the set of estimated coefficients in (18) for the X variable are not statistically different from zero ($\sum d_i = 0$);

Unilateral causality from X to Y, when all the lagged Y variable to the equation (17) are not statistically different from zero ($\sum a_i = 0$) and the set of coefficients for the lagged variable X in (18) are statistically different from zero ($\sum d_i \neq 0$);

Bicausality or simultaneity: when the sets of lagged coefficients of X and Y are statistically different from zero in both regressions;

Independence: When in both regressions, the sets of lagged coefficients of X and Y are not statistically different from zero.

DESCRIPTION OF THE DATA

In this work we use a series of yield and harvested area of soybeans in the state of Parana released by the Brazilian Institute of Geography and Statistics – IBGE from April 1998 to August 2011. The data set of prices is released by the Department of Rural Economy of the Agriculture and Food Supply of state of Parana - DERAL. The price is given in 60-kilo bags of soybean and prices were deflated to August 2010 by the General Price Index - (IGP-DI) of the Getúlio Vargas Foundation - FGV (2011).

RESULTS

Figures 1 and 2 shows the evolution of the variables under study during the analyzed period. One should note that, while the yield of soybeans in Paraná increased 34% from April 1998 to August 2011, the planted area increased 64.3% over the same period. On the other hand, prices on August 2011 were only 13% above prices on April 1998. The area occupied by the soybean in the state of Parana increased throughout the period, the average yield in kg / ha (expected yield) decreased significantly in 2004 returning to recover only in 2006 and in 2009 has declined. The prices behavior in the spot market oscillated throughout the period, with a peak value between 2002 and 2004, and on October 2002 prices doubled the prices observed in April 1998. At the beginning of 2004 these prices started a downward trend which changed only in 2007. In 2009, due to the international economic crisis prices decreased and at the end of the period, in August 2011, prices were 13% above April 1998 prices.

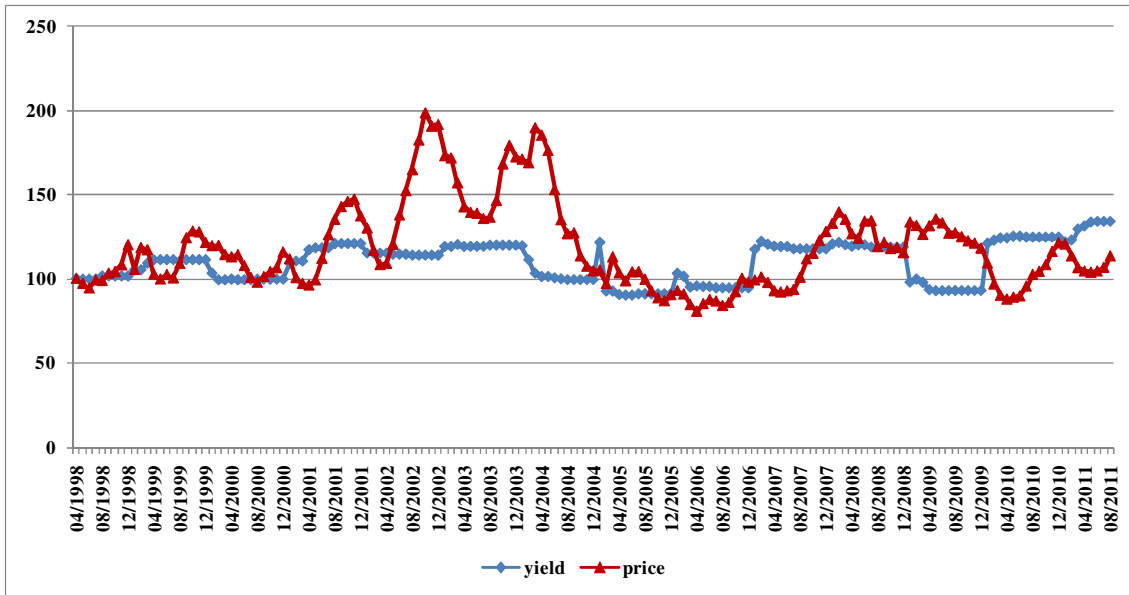


Figure 1 – Evolution of the physical price and expected yield of soybeans in Parana.

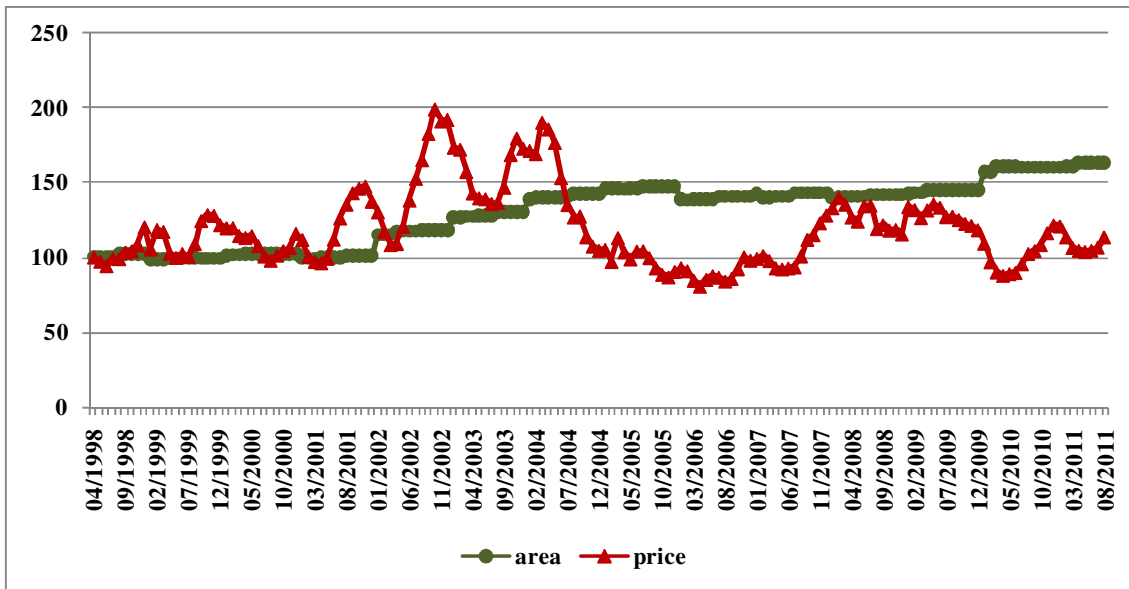


Figure 2 – Evolution of the physical price and area with soybeans in Parana.

We transformed the data using the logarithmic transformation in order to stabilize the variance of the data. Another result of working with the logarithmic

transformation of the series is that the transformed value refers to the rate of growth. Thus, we obtain directly the elasticities.

To test the stationarity of the three series (logarithmic transformation) we used the Augmented Dickey-Fuller test. The number of lags were determined by AIC and SBC criteria and analysis of auto correlation and partial auto-correlation. The ADF test results - Table 1 indicated that the three series are difference stationary at the level of 1%.

Table 1- ADF test for área, yield, and price

MODEL	Test Statistic	Critical Values	area	yield	Price	Reject
		1%				
$\Delta x_t = \alpha + \beta t + \rho x_{t-1} +$ Q (1 lag)	τ_τ	-4,04	-1,59	-2,31	-3,03	Not
	$\tau_{\beta\tau}$	3,53				Not
	ϕ_3	8,73	1,27 ***	2,76 ***	4,63 ***	Not
$\Delta x_t = \alpha + \rho x_{t-1} + \varepsilon_t$	τ_μ	-3,51	-0,66	-2,19	-2,97	Not
	$\tau_{\alpha\mu}$	322				Not
	ϕ_1	6,7	2,91	2,56	4,44	Not
$\Delta x_t = \rho x_{t-1} + \varepsilon_t$	τ	-2,6	2,31	0,53	0,02	Not
$\Delta\Delta x_t$	τ	-2,6	-8,41	-9,21	-7,07	Yes

Fonte: Data from work.

*** Q Ljung-Box Q-Statistics significance.

To determine the temporal relationships of precedence among the three variables we applied the Granger causality test. As a result of this test, two significant relationships at the 5% significance level with two lags were obtained. Thus, in terms of temporal precedence, variable expected yield is important to help explain the behavior of the spot market price of soybeans in Parana, the other relationship - price on area - are also important to help explaining the decision to increase or not the soybean area based on the price.

The cross-correlation test corroborated the Granger causality test. The analysis of price and yield indicated that there is no significant relationship in the contemporary period, but a significant relationship of yield explaining price with a lag period. Cross-correlation between price and area did not show significant contemporaneous relationship, this was only significant one month lag in the sense of spot market price explaining area. Considering the area and expected soybean yield significant relationships were not found neither in the contemporary period nor in the lags.

Tabela 2 – Johansen cointegration test between yield, area and price

Null Hypothesis	Alternative Hypothesis	Estatistic Trace	Critical values - 5%	Critical values - 1%
$r \leq 2$	$r > 2$	5,33	12,25	16,26
$r \leq 1$	$r > 1$	18,71	25,32	30,45
$r = 0$	$r > 0$	42,18	42,44	48,45

Fonte: Data from work.

To set the model to be used, VAR or VEC, we applied the Johansen cointegration test between the three series (1) - Table 2. It is observed that, at 5%, we accept the hypothesis that the rank of the matrix is zero, so there is no cointegration and the VAR model is appropriate to the time series in first difference.

The equation for growth rate of the VAR model indicates that only de expected yield and price (5% significance), both in the first time difference, were important to predict the prices behavior on the spot market. Now, looking at the yield any variable was significant to explain their behavior. Considering the area, the constant of the model and the first time difference of the prices are important to explain the growth of the area in the state at 5% level of significance.

CONCLUSION

The objective of this study was to evaluate the impact of changes in planted area and soybeans yield in the state of Parana on the spot price. In other words, we analyze the impact of changes in yield and expectations of the planted area on prices and anticipate possible losses that may occur or anticipate possible adverse reactions of the economic agents. To evaluate the interaction between the three chosen variables and study the effects expected in the course of time, unit root tests, cointegration and causality were performed. The results of these tests, as well as the results of the estimates of the VAR model (assuming a significance level of 5%) indicated that to predict the rate of growth of prices in the spot market only expected yield and price, both in first difference time were significant. Regarding the yield, neither the rate of growth of prices nor the expectation of yield were significant in explaining their behavior. For the rate of growth of the area with soybeans in Parana, the constant of the model and the first difference of the price are important to explain his behavior.

We conclude, therefore, that prices react to information on crop yield with a lag period, since the yield data are reported one month late. In the case of the area, it seems that prices are an important variable when deciding whether or not to increase the area however, we considered the price in the previous month.

It is important to emphasize that the objective of this study was to verify the relationship between prices and the variables that determine the supply of the product (area and yield) and not explain its behavior, since to fully explain its behavior would be necessary include the demand side.

5. REFERENCES

ALVES, L.R.A. Price transmission between products of this sector of the State of Sao

Paulo. 2002. 107 p. Dissertation (MSc in Applied Economics) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2002.

ENDERS, W. Applied Econometric Time Series. New York: John Wiley & Sons, 2004, 2nd ed., P. 466

FULLER, WA Introduction to Statistical Time Series. New York: John Wiley & Sons, 1976. 424 p.

GRANGER, C.W.J. Investigating causal relations by Econometric models and cross-spectral methods. *Econometrica*, Menasha, v. 37, n. 3, p. 424-438, Aug.1969.

HAMILTON, J. D. Time series analysis. New Jersey: Princeton University Press, 1994. 799p.

Brazilian Institute of Geography and Statistics - IBGE. Systematic Survey of Agricultural Production - LSPA. Available at:
<http://www.ibge.gov.br/home/estatistica/indicadores/agropecuaria/lspa/>>. Accessed 12 August 2011.

JOHANSEN, S. Statistical Analysis of cointegration vectors. *Journal of Economics Dynamics and Control*. v.12, n.2 / 3, pg. 231-254, 1988.

OZAKI, V.A. Spatial analysis of agricultural productivity in the state of Paraná: implications for crop insurance. *Journal of Economics and Rural Sociology*, v. 46, p.869-886, 2008.