HELLENIC OPEN UNIVERSITY

SCHOOL OF SOCIAL SCIENCES

MASTER IN BUSINESS ADMINISTRATION (MBA)

Postgraduate Dissertation

Cotton price forecasting using univariate and multivariate time series models

Nikolaos Katsenios

Supervisor: Prof. Constantinos Katrakilidis

Patras, Greece, June 2021

Cotton price forecasting using univariate and multivariate time series models

Nikolaos Katsenios

Supervising Committee

Supervisor:	Co-Supervisor:
Constantinos Katrakilidis	Stavros Ntegiannakis
Professor HOU	Professor HOU

Patras, Greece, June 2021

Acknowledgments

First of all, I would like to express my gratitude to my Supervisor, Professor Constantinos Katrakilidis for his guidance and support throughout this study. His teaching, patience and encouragement were of high importance for the completion of this study. I would also like to thank my co-Supervisor, Professor Stavros Ntegiannakis for his help.

Finally, I would like to thank my wife Spyridoula for her assist and patience and my kids Charalampos and Lefkothea for their support.

Dedicated to

my son Charalampos

and my daughter Lefkothea

Abstract

This dissertation investigates the dynamic relationships among cotton price, crude oil price and world cotton stocks to world cotton use ratio. The time series data that we use in this study were retrieved from international influential organizations, like the United States Department of Agriculture (USDA) and the World Bank. The collection of the above-mentioned secondary data covers a period of more than 30 years (365 monthly observations). The empirical analysis is based on the Johansen cointegration methodology. The results of the analysis indicate that there is a significant long-run causal effect that is directed towards the cotton prices and another long-run causal effect that is directed towards the stocks to use ratio. According to the Granger causality test, we identify a short-run causal effect from cotton price to crude oil price. Impulse response and variance decomposition analysis applied to our results to study the out of sample forecasting behavior of the system. The results indicate that a shock at the value of world cotton stocks to world cotton use ratio is starting to reduce the price of cotton approximately 4 months later and maximizes about 6 months later. A shock at the price of crude oil leads to a mild increase of the cotton price during a 24-month period. As for the explanatory power of the variables, cotton stocks to use ratio explains only up to 20% of the behavior of cotton prices. The use of Johansen cointegration test, which is a methodology widely accepted for its reliable results contributes to the econometrics science with novel results, regarding the ability to make models that can forecast the movements of cotton price in the long-run and in the short-run.

Keywords

Cotton prices, cointegration, causality, agricultural commodities

Περίληψη

Αυτή η διατριβή διερεύνα τις δυναμικές σχέσεις μεταξύ της τιμής του βαμβακιού, της τιμής του αργού πετρελαίου και της αναλογίας των παγκόσμιων αποθεμάτων βαμβακιού προς την παγκόσμια χρήση. Τα δεδομένα των χρονοσειρών που χρησιμοποιούνται σε αυτήν τη μελέτη ανακτήθηκαν από έγκυρους διεθνείς οργανισμούς, όπως το Υπουργείο Γεωργία των ΗΠΑ (USDA) και η Παγκόσμια Τράπεζα. Η συλλογή των προαναφερθέντων δευτερογενών δεδομένων καλύπτει μια περίοδο άνω των 30 ετών (365 μηνιαίες παρατηρήσεις). Η εμπειρική ανάλυση βασίζεται στη μεθοδολογία της συνολοκλήρωσης Johansen. Τα αποτελέσματα της ανάλυσης δείχνουν ότι υπάρχει μια σημαντική μακροπρόθεσμη αιτιώδης επίδραση που κατευθύνεται προς τις τιμές του βαμβακιού και μια άλλη μακροπρόθεσμη αιτιώδης επίδραση που κατευθύνεται προς την αναλογία αποθεμάτων βαμβακιού προς την χρήση. Σύμφωνα με τη δοκιμασία αιτιότητας κατά Granger, εντοπίστηκε μια βραχυπρόθεσμη αιτιώδης επίδραση από την τιμή του βαμβακιού προς την τιμή του αργού πετρελαίου. Η ανάλυση της συνάρτησης αιφνίδιων αντιδράσεων και η ανάλυση διάσπασης της διακύμανσης εφαρμόζονται στα αποτελέσματά μας για να μελετήσουμε τη συμπεριφορά πρόβλεψης του συστήματος. Τα αποτελέσματα έδειξαν ότι ένα σοκ στην τιμή της αναλογίας των παγκόσμιων αποθεμάτων βαμβακιού προς την παγκόσμια χρήση βαμβακιού αρχίζει να προκαλεί μείωση στην τιμή του βαμβακιού περίπου 4 μήνες αργότερα και μεγιστοποιείται περίπου 6 μήνες μετά. Επίσης, ένα σοκ στην τιμή του αργού πετρελαίου οδηγεί σε ήπια αύξηση της τιμής του βαμβακιού κατά τη διάρκεια μιας περιόδου 24 μηνών. Όσον αφορά την επεξηγηματική ισχύ των μεταβλητών, η αναλογία αποθεμάτων προς τη χρήση του βαμβακιού εξηγεί μόνο έως 20% της συμπεριφοράς των τιμών του βαμβακιού. Η χρήση της μεθόδου του Johansen, η οποία είναι μια μέθοδος ευρέως αποδεκτή για τα αξιόπιστα αποτελέσματά της, συνεισφέρει στην επιστήμη της οικονομετρίας με νέα αποτελέσματα, σχετικά με την ικανότητα δημιουργίας μοντέλων που μπορούν να προβλέψουν τις μεταβολές της τιμής του βαμβακιού σε μακροπρόθεσμο και βραχυπρόθεσμο ορίζοντα.

Λέξεις – Κλειδιά

Τιμές βαμβακιού, συνολοκλήρωση, αιτιότητα, αγροτικά εμπορεύματα

Table of Contents

List of Figures

List of Tables

List of Abbreviations & Acronyms

ADF	Augmented Dicker Fuller
ADF-GLS	Augmented Dickey-Fuller Generalized Least Squares
CAI	Cotton A Index
CE	Cointegrating Equation
CME	Chicago Mercantile Exchange
COIL	Crude Oil
CZCE	Zhengzhou Commodity Exchange
ECM	Error Correction Model
FAO	Food and Agriculture Organization
GDP	Gross Domestic Product
ICAC	International Cotton Advisory Committee
ICE	Intercontinental Exchange
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
LS	Least Squares
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
STU	Stocks To Use
US	United States
USDA	United States Department of Agriculture
VAR	Vector Auto Regression
VECM	Vector Error Correction Model
WTO	World Trade Organization
ARMA	Autoregressive Moving Average
NLS	Nonlinear Least Squares

1. Introduction

In recent decades, agricultural commodity prices, including cotton price, have shown high variability. These fluctuations are the result of changes at the fundamental data of the market (demand, production, etc.), as well as the outcome of some economic events and changes at the policies of the governments.

Worldwide, there is an increasing demand from the stakeholders of the agricultural commodities sector for models that can forecast the movements of agricultural products prices in the long-run and in the short-run. These models are based on a wide variety of variables, in an effort to develop reliable predictions for their relationships. Farmers are willing to use such forecasting models, in order to plan their sales of their production, as a tool to reduce uncertainty. Presently, the use of models is very limited, as their performance in the longer horizons is not so trustworthy. The use of advanced methodologies that are widely accepted for their reliability could contribute to the creation of models that can forecast the movements of cotton price in the long-run and in the short-run.

Among agricultural commodities, cotton has been studied to a lesser extent compared to other commodities, regarding the dynamic relationships among cotton price and a wide variety of variables. However, cotton is a very important commodity worldwide and the results of such research are valuable for all stakeholders of the sector.

Aim of the study

The aim of this study is to investigate the ability to use a heterogeneous variable (world production, consumption and stocks) in order to create price-forecasting models for cotton. The data were retrieved from international influential organizations, like USDA and the World Bank. The collection of the above mentioned secondary data covers a period of more than 30 years (365 monthly observations). The forecast model that will be created is the result of univariate and multivariate time series models. To the best of our knowledge, till now, cotton behaviour forecasts were obtained by using either univariate models or multivariate models or techniques based on single equation estimations. However, there is no a research that combines more possible deterministic factors under a system of equations and through this system to explore the possible long run and short run dynamics of the involved variables. Considering this observation, we attempt to cover this gap by using the appropriate cointegration technique suggested by Johansen. The use of Johansen cointegration test, which is a methodology widely accepted for its reliable results contributes to the econometrics science with novel results, regarding the ability to make models that can forecast the movements of cotton price in the long-run and in the short-run.

The structure of this thesis is as follows: Chapter 2 provides some important information for cotton, while Chapter 3 presents the relevant literature review. In Chapter 4 the methodological framework is presented and the empirical results are in Chapter 5. Finally, Chapter 6 summarizes and concludes.

2. Cotton

Cotton (*Gossypium hirsutum*) is a profitable industrial plant that is cultivated in more than 75 countries worldwide, both in the Northern and Southern hemispheres for its fiber and seed. Almost 70% of world production is derived from irrigated condition (ICAC, 2016). Especially for the developing countries it is considered as a major cash crop and it is often mentioned as the white gold, since cotton trade brings foreign exchange (Khan et al., 2020). Cotton yields have been stabilized during the last years, as a result of pest outbreaks, water shortage and because countries with low yields have managed to follow up and close the gap from the top yielding countries (OECD et al., 2020). The top producing countries are India (23%), China (21%), USA (15%), Brazil (10%) and Pakistan (7%), accounting for approximately 75% of world cotton production (Figures 1 and 2).

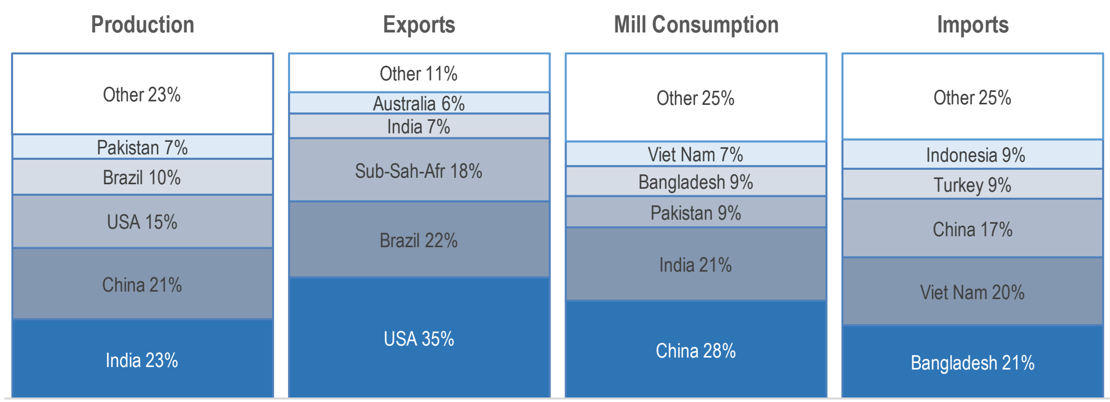

Figure 1. Share of world totals of production, exports, mill consumption and imports of cotton
Source: OECD-FAO Agricultural Outlook 2020-2029 (OECD et al., 2020).

World cotton use has a clear trend of stable increase during the last years. More specifically, during the period of the marketing year 1995-96 till marketing year of 2017-18 has increased by 70.10%. Actually, it is estimated that the international purchases of the cotton during 2017-18 were approximately 50 billion dollars, and about 65% of the total quantity sold was imported from Asian countries (Khan et al., 2020).

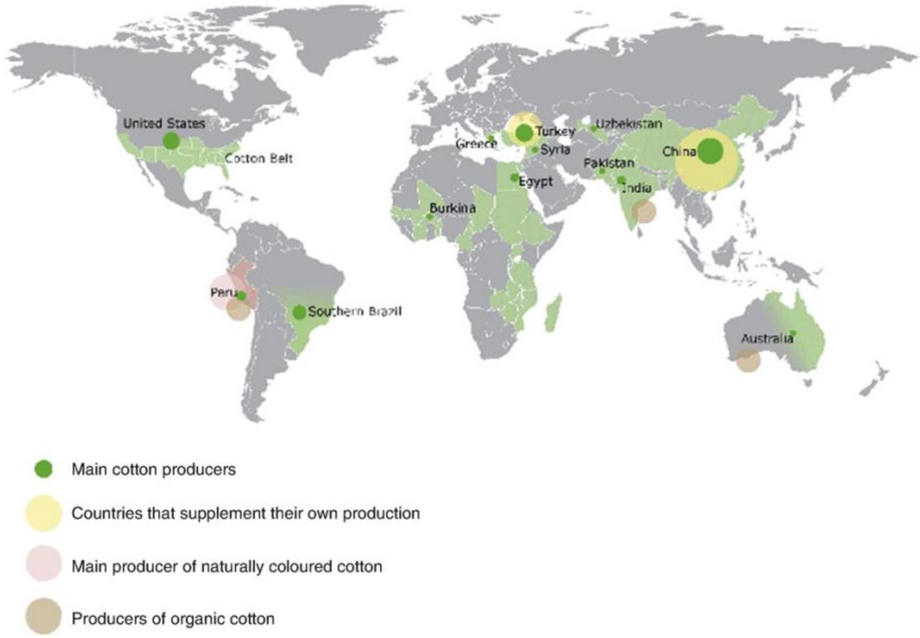

Figure 2. The main cotton producing countries of the world (Khan et al., 2020)

2.1 Cotton history and uses

The exact origin of cotton is still undefined. It is cultivated for its fiber since ancient times, and it is indigenous at areas with tropic climate in Asia and Africa. Excavations in India have uncovered remains of cotton cloth dating to around 3000 BC. It first came to Greece from Asia during the time of Alexander the Great around 325 BC. Its cultivation then spread to other European Mediterranean countries. In those years, cotton was referred to as a tree, which proves that they grew arboreal varieties of cotton. The cultivation of cotton in Greece is mentioned by Pausanias in 2nd century AD with the name "βύσσος". Its cultivation expanded on a large scale around 550 AD.

2.2 Cotton as a commodity

Cotton belongs to the group of soft commodities, a term that is used to separate the commodities that are usually grown, instead of the commodities that are mined and are called hard commodities. The price of soft commodities is determined by the fundamental data of the market (supply and demand) more intensively than other products. Cotton has been traded at the New York Board of Trade since 1870, which from 2007 was acquired from Intercontinental Exchange (ICE). Its transactions are made following some contract specifications (Table 1). Cotton is also traded at the Chicago Mercantile Exchange (CME) with similar contract specifications. Since June 2004, Zhengzhou Commodity Exchange (CZCE) of China, launched a future contract for cotton (Cotton #1), where large volumes of the product are traded.

Farmers participate at the market of cotton future contracts in order to hedge the risk and investors to earn money. The market activity of cotton has some clear signs of seasonality, a characteristic that is common for the agricultural commodities because of the determined growing periods that are based on the four seasons (Olen & Andersson, 2013).

Table 1. Market specifications of Cotton No.2 futures at Intercontinental Exchange (www.theice.com)

Trading Screen Product Name	Cotton No. 2 Futures
Trading Screen Hub Name	NYCC
Contract Symbol	CT
Contract Size	50,000 pounds net weight
Quotation	Cents and hundredths of a cent per pound
Contract Series	March, May, July, October, December
Minimum Price Fluctuation	1/100 of a cent (one "point") per pound equivalent to $5.00 per contract.
Settlement	Physical Delivery
Daily Price Limit	Futures contracts are subject to a daily price limit that can range from 3 to 7 cents per pound.
Deliverable Origins	US Origin only.
Delivery Locations	Galveston, TX, Houston, TX, Dallas/Ft. Worth, TX, Memphis, TN and Greenville/Spartanburg, SC.
Grade/Standards/Quality	Quality : Strict Low Middling Staple Length: 1 2/32nd inc

Cotlook A index

Cotlook A index is the main reference for the international cotton prices. The Cotlook A index is published since 1966 from Cotton Outlook and it intends to be representative to the international raw cotton market prices. This index is in fact the average price of the five lowest prices of eighteen selected quotations in a daily basis (Figure 3). For the uniformity and representativeness of the index, the calculations are made on sales of cotton with base quality MIDDLING 1-1/8″ (www.cotlook.com). The procedure of the creation of this index is dynamic and it has changes throughout the more than fifty years of life, in order to remain the primary indicator of the international cotton prices. A index is a widely accepted price index and it is representative for cotton prices worldwide (Figure 4).

Australian	Greek	Mexican
Benin BELA*	Indian medium grade **	Paraguayan
Brazilian	Ivory Coast BEMA*	Syrian
Burkina Faso RUDY*	Mali KATY*	TanzanianType 1 SG
California/Arizona	Memphis/Eastern	Turkish S. Eastern Std 1 RG
Chinese 328	Memphis/Orleans/Texas	Uzbekistan

* Only two African Franc Zone origins are currently allowed to figure in the A Index calculation on any day
** Applicable ICS standard, as adopted by the ICA is ICS 105 Fine

Figure 3. The 18 selected quotations that contribute to the calculation of Cotlook A index (source: www.cotlook.com)

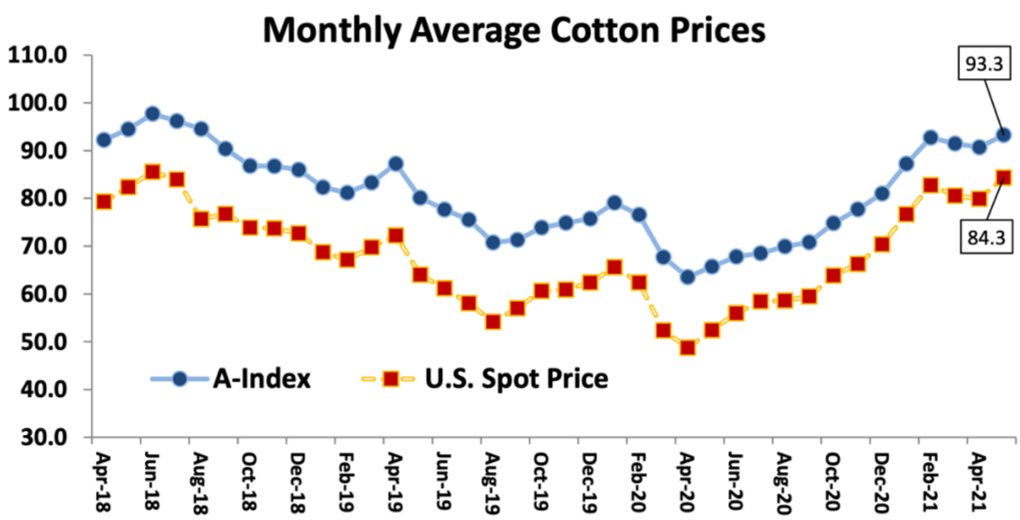

Figure 4. Monthly average cotton prices
Source: www.usda.gov. Cotton: World Markets and Trade: Global 2021/22 Cotton Consumption Highest in 4 Years (May, 2021)

The main substitute material for cotton for the textile industry is the polyester. In the last years, there is a stabilization of the ratio of cotton and polyester prices. One reason for this stability is the turn of China to a more green economy that slows down the polyester production (OECD et al., 2020).

2.3 Factors that influence world cotton price

China is the largest consumer accounting for 28% of the total cotton use for 2020 marketing year (Figure 5), while Bangladesh, Turkey and Vietnam maintain a strong growth of demand to supply their spinning and textile industry (OECD et al., 2020). For a long period, China's demand for cotton was constantly increasing, until 2007 when it peaked. After this year, a decline at the demand of cotton started, because of the increased labor costs and government regulations in China, a situation that made industry to move to other Asian countries (OECD et al., 2020).

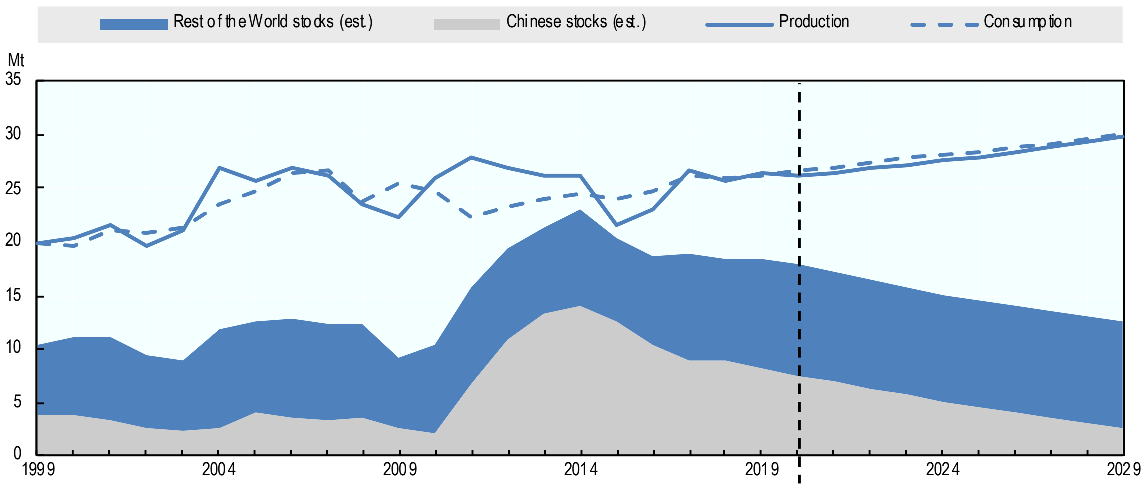

Figure 5. World cotton stocks in China and rest of the world, production and consumption
Source: OECD-FAO Agricultural Outlook 2020-2029 (OECD et al., 2020).

2.4 Cotton price forecasting

There are many uncertainties that have major or minor effects at the forecasting of commodities and especially agricultural commodities such cotton. The COVID-19 pandemic for example, reminded how volatile can the markets become because of a single event, not including war conflicts (Figure 6). In addition, climate change (intense weather phenomena) is a major factor that intensifies the uncertainty of the cotton market. Cotton market is considered as sensitive to external shocks that produce high fluctuations on the price. A typical example is the significant increase in cotton prices in the marketing year 2010/11, where Cotlook A index increased up to 5,06 $/kg (Figure 7). This fluctuation was a result of multiple factors, as the increase of oil and polyester prices and a surprising high demand of cotton from China because of the decision of its government to increase its stockpiles (OECD et al., 2020).

There is a desire at the international cotton market to create a model that is capable to predict reliably the price short-term and if possible long-term. In USA, for many years (1929-2008) there was a legislation the forbid the USDA to publish price forecasts for cotton (Isengildina-Massa & MacDonald, 2009). After the restart of USDA cotton price forecasts, the evaluation of the predicted prices showed that there was a tendency of the model to overestimate the cotton prices and this was a result of some structural changes in the industry (Isengildina-Massa & MacDonald, 2009).

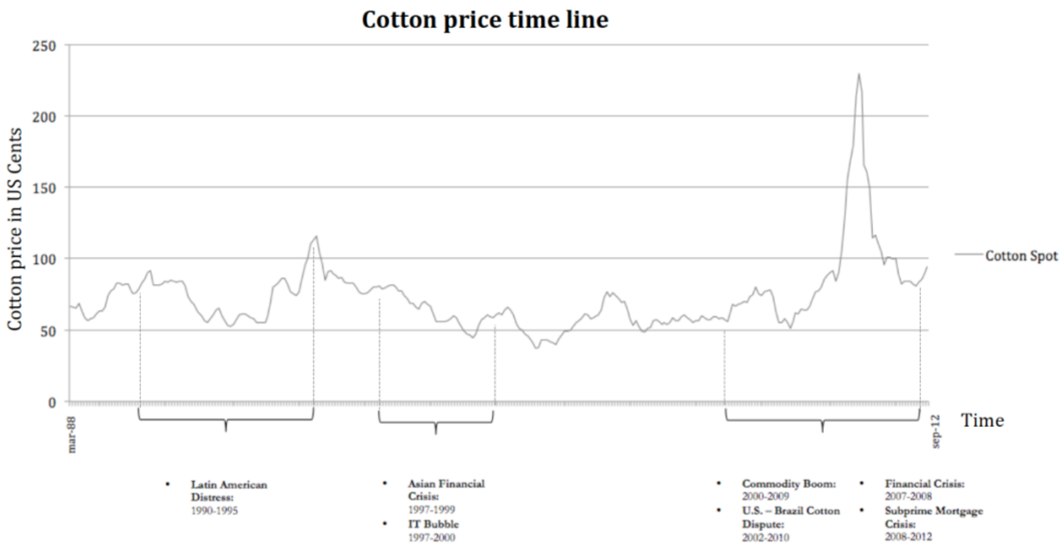

Figure 6. The effect of market shocks at the price of cotton since 1988 (Olen & Andersson, 2013)

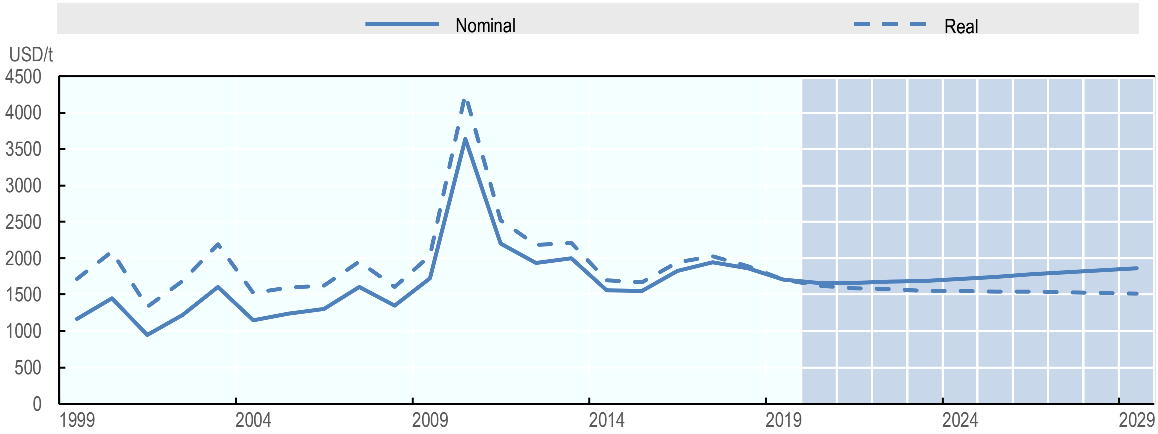

Figure 7. Nominal and real price of Cotlook A index in usd/t (average of marketing year August-July)
Source: OECD-FAO Agricultural Outlook 2020-2029 (OECD et al., 2020).

2.5 Role of crude oil price in cotton production chain

Crude oil, known also as petroleum, is a product of geochemical processes that converted fossilized organic materials to oil. In forecasting studies of agricultural products, oil is usually a major influential parameter, as it affects the production chain from many aspects (Olen & Andersson, 2013). The production cost of cotton cultivation is affected from oil prices, as oil is the fuel used by tractors for many cultivation operations, such as tillage, sowing, application of pesticides and herbicides, harvest, etc. Recent research indicated that there is an increase in the dependence of crude oil price and cotton according to data of price series from 2002 to 2014 (Ghorbel et al., 2017).

2.6 Stockpiles and consumption of cotton worldwide

Cotton is the raw material that is used from the mills to produce yarn. The first step is the removal of the seeds from the fibers, a procedure that is known as ginning. After this, follows the procedure of spinning, where the fibers are used to make yarns. Then the yarns are used to make fabrics. When we talk about cotton consumption, in the commodities market, we refer to the use of cotton fibers from mills than make yarn. The use of cotton from mills is depended on the world demand for textiles. World cotton consumption per capita has a stabilizing trend during the last years and this is a result of increased competition from its main substitute that is the synthetic fibers. According to the latest data, world cotton consumption presented an all-time record at 27 Mt during 2007, and then decreased slightly at 26 Mt for the marketing period of 2017-19 (OECD et al., 2020). The increase of income in many developing countries, is expected to lead to an increase of approximately 1.3% per year in the next five years.

The accumulation of cotton stockpiles is usually a government decision that is related with the choice of the government to strengthen the economic activity related to the cloth industry. An example is the government purchases of cotton from China during 2000-2010. The increase in labor cost resulted in the change of the government's policy and a transport for the mill and cloth industry from China to other Asian countries with lower labor cost, such as Vietnam and Bangladesh.

2.7 Events that shocked the international cotton market

According to the available historical data of cotton prices, the events that shocked the cotton market to a large or a smaller extent, have increased during the recent decades. These dates are considered as structural breaks, which means that are unexpected changes in the time series that could lead to forecasting errors.

In a research that tried to identify the structural breaks in the commodities future markets for the period 1990-2009 using Bai and Perron test, the findings of the analysis indicated that regarding cotton market, one break was identified at 8 March 2001, (Coakley et al., 2011). Some cycles of extreme cotton price changes that have been identified from a USDA report analysis of the monthly prices of more than 200 years, are the periods 1986-87, 1992, 2001-05 and 2009-12 (MacDonald & Leslie, 2018). At similar research including agricultural commodities such as maize, cotton, sugar, etc., it is usual to identify structural breaks at the examined time-series data, that at most of the cases are connected to some major events, that are related with government decisions, legislation changes, and the changes at the prices of other competitive or complementary commodities.

Briefly, some events that could have affected the cotton market is the Latin America Distress (early 1990s), the Asian Crisis (1999), the Dot-Com bubble (late 1990s), the 2008 Commodity Boom and the Brazil Cotton dispute of 2002-2010 (Olen & Andersson, 2013). At a research for US corn for ethanol use, the identified break date falls in the period when public discussions started for the new legislation regarding the establishment of a Renewable Fuel Standard program (Oladosu et al., 2021). In a cotton study, with daily prices, a breakpoint was identified at mid-February, 2010, while a second was found at 1 December, 2003 (Olen & Andersson, 2013).

The identification of structural breaks in the vulnerable agricultural market and more specifically for cotton prices could contribute as a useful tool in the decision-making progress of the development of forecasting models (Katrakilidis et al., 2005). As a possible explanation for the increased market prices in 2008, many researchers claim that this rise was not driven by economic fundamentals but rather by speculation (Power & Robinson, 2009). The commodity price boom that took place in 2008 affected in general the prices of the commodities. Especially ethanol, that is a substitute of crude oil, and a competitive (in areas) cultivation of cotton has an increasing importance at the cotton market. Structural break analysis has been used on quarterly data of US corn use for ethanol for years 1986 to 2017 and identified three breaks because of the transition from the period before 2000 when there was slow growth, to after 2000 when use increased rapidly (Oladosu et al., 2021).

3. Literature Review

Cotton forecasting models are based on a wide variety of variables. In a recent study, the variables were cotton price, cotton Futures, GDP total for OECD countries, maize, oil, S&P 500, sugar and wool in order to make a model to predict in high accuracy the future cotton price (Olen & Andersson, 2013). In another research, they used US exchange rate index, world demand, supply and stocks data, in order to predict the Cotton A index (MacDonald, 2006). An important feature of the world cotton market that affects the cotton price is the ending stocks of each marketing year and more specifically, where are they and what is the relationship, and the ratio compared to the use. Alternative measures of the above-mentioned data are usually used in order to make more accurate predictions. Variables such as: World stocks-to-use, World minus China stocks divided by world minus China consumption, World minus China stocks divided by world consumption, World use-to-stocks, World minus China consumption divided by world minus China stocks, World consumption divided by world minus China stocks have been used (MacDonald, 2006). Other researchers prefer to include data related to the income and the expenses of the farmers, where oil plays a crucial role as a major expenditure (Chen & Bessler, 1990). Oil price has been found that contributes to the volatile clustering and jump behavior of the agricultural commodities, including cotton (Zhang & Qu, 2015). In a research published from USDA officials, the variables that were used to predict the price of cotton were U.S. cotton supply, U.S. stocks-to-use ratio, China's net imports as a share of world consumption, the proportion of U.S. cotton in the loan program, and the world supply of cotton (Isengildina-Massa & MacDonald, 2009). In a study that aimed to forecast the daily futures price of cotton of the ICE exchange, they used the daily cotton certified stocks data, as well as an annual series of US plantings acreage, monthly data on U.S. net exports, US ending stocks, US total use and US stocks-to-use ratio (Power & Robinson, 2009).

In a model comparing research from International Monetary Fund (IMF), that examined 15 commodities, specifically for cotton found that the most suitable model for one quarter horizon was that which took the form of a unit root with futures prices as an additional exogenous model (Husain & Bowman, 2004).

Usually unit-root test are conducted in order to identify the stationarity of the variables. In a recent research, that was based on cotton and crude oil price, the results indicated that these variables, among others, were non-stationary (Olen & Andersson, 2013). Especially Cotton A index non-stationarity has been confirmed by many researchers (MacDonald, 2006). Moreover, time series data consisted of world stocks, stocks to use ratios, world stocks minus China stocks, etc have been found to be non-stationary at levels (MacDonald, 2006). In a research that used daily data of cotton price for a decade (2004-2014) in China, the stationarity tests indicated that cotton price was not stationary at the level and was stationary at the first difference (Zhang & Qu, 2015). In a recent research, unit root tests Augmented Dicker Fuller (ADF) and KPSS were used to understand the stationarity of the variables, prior the implementation of the causality tests (Oladosu et al., 2021).

There is an increasing demand for cotton price forecasting models. Even the ban that was active in the USA from 1929 due to a Congress legislation, was removed at 2008 and USDA relaunched its publications related to cotton price forecasting (Isengildina-Massa & MacDonald, 2009). In fact, the use of such forecasting models from the farmers as a tool to reduce uncertainty is very limited, as their performance in the longer horizons is not so trustworthy (Olen & Andersson, 2013). Forecasting failures for cotton price are attributed in a great degree at the unstable behavior of China's stockpiling program (MacDonald, 2006). It is generally accepted that the existing cotton price forecasting models present poor predictive capability and for this reason there are many efforts to review and improve the existing models (Isengildina-Massa & MacDonald, 2009).

Most of the commodity price forecasting studies, that include the prediction of cotton price, agree that the futures prices can provide reliable results about possible developments in spot prices over the longer term, at least in directional terms, performing a high reliability result especially at the two-year horizon (Husain & Bowman, 2004). A model developed by USDA officials claims that can explain 68% of the variation of the cotton price in the USA based on data from 1974 to 2007 (Isengildina-Massa & MacDonald, 2009). The same model predicts the price of cotton with a root mean squared error 6 cents/pound, which is about 10% of the mean price of the period 1974-2007. This increased error implies that may be another significant variable (i.e., polyester price) that is not used in this model, that can be added in the future and improve the reliability of the results. Another factor that can be influencing at the forecast errors that occur, is the systematic errors that are conducted during the procedure of the collection of data that are used as dependent variables, such as the world supply and demand (Isengildina-Massa & MacDonald, 2009). The data of agricultural commodities are characterized by both long memory and structural breaks, which means that the persistence could lead to an improvement of the forecasting models, but the multiple structural breaks intensify this challenge (Coakley et al., 2011). In the forthcoming period until 2027, a USDA report claims that the volatility will be likely greater than the period 2016-17, when the instability was unusually low because of the decision of China to reduce its stocks from the National Reserves (MacDonald & Leslie, 2018).

4. Methodological framework

4.1 Stationarity – Unit Root tests

At this study we used the Dickey-Fuller GLS test and the Kwiatkowski-Phillips-Schmidt-Shin test complementary to investigate the stationarity of our variables. It is important to note that the null hypothesis of KPSS test is that the variable is stationary while the alternative hypothesis is that is not stationary. On the contrary, Dickey-Fuller GLS test has a null hypothesis that we have a unit root, while the alternative hypothesis is that the time series is stationary.

Time series is a series of data points in successive order over a period of time with stable intervals. Variables that change over time, like for example price of stocks, commodity products can create time series. The historical values of such variables can be used for time series forecasting methods and predict the future activity. Data that are non-stationary are in general unpredictable and cannot be used to make forecasting models. For the forecasting methods that are used at the econometrics, the most common assumption is that the data are stationary. A time series it is said that is stationary when its statistical properties such as mean, variance and autocorrelation are stable over time.

To test the Stationarity of a time series, we usually use unit root tests. By the term unit root in macroeconomic studies, we mean that some root of the polynomial below is equal to the unit.

$$f(x) = 1 - \rho_1 x = \rho_2 x_2 - \rho_3 x_3 - \ldots - \rho_n x_n = 0 \qquad (4.1)$$

In this case any exogenous change on an endogenous macroeconomic variable can have a permanent effect on it. This result can be obtained from a first order autoregressive model AR (1) with an autocorrelation coefficient close to the unit and the white noise u_t playing the role of the random variable.

$$Y_t = \rho Y_{t-1} + u_t \qquad (4.2)$$

The stationarity of the time series, that is the non-existence of a unit root, is checked based on the assumptions:

H_0: $p \geq 1$, the time series is non-stationary, there is a unit root

H_1: $p < 1$, the time series is stationary, there is no unit root

Some popular test unit root tests are the Augmented Dickey-Fuller test, Phillips-Perron test, KPSS test, ADF-GLS test, Breusch-Godfrey test, Ljung-Box test and Durbin-Watson test. A stationary series is said to be integrated of order zero, *I* (0), while a non-stationary series has a higher order of integration. The series that is differenced *n* times in order to become stationary, is said to be integrated of order *n*, *I* (n). The majority of economic time series are stationary or becoming stationary at the first differences.

Stationarity is very important condition as it means that the parameters of the model remain stable and it is applicable throughout time (Oladosu et al., 2021). Another important factor that makes the the unit-root test significant is the fact that if the results of the unit-root test present non-stationarity in levels and stationarity in first differences, then the results of cointegration tests are valid (Dimitriadis & Katrakilidis, 2020).

4.2 Cointegration

Cointegration is a method based on the synchronization of non-stationary time series that is able to identify the relationship between two or more variables in the long term. There are two basic methods that are used to test the cointegration of two or more variables. These methods are used to identify the relationship in the long run of two or more time series. The first one is the Engle-Granger two-step method, that produces a single equation model. The second one is the Johansen test, that allows more than one cointegrating relationships. In our study we use the Johansen methodology, which produces a system of equations.

Usually, the economic time series present a trend. Taking the first differences, we have loss of long-term properties, and for this reason we are looking for a specific model that will combine both the long-term and short-term characteristics of the examined variable and at the same time will maintain the property of stationarity. The concept of cointegration allows us to describe the equilibrium relationship, that is a long-run relationship between two or more (economic) variables, each one characterized by stationarity.

The concept of cointegration is based on the synchronization of non-stationary time series. That is, if two or more variables move in the same direction in the long run, there is likely to be a long-run equilibrium relationship between the variables, something that is not valid in the short run. Then the regression results may not be fictitious and therefore the

conclusions based on t and F statistics may be valid. In the short run, the variables may follow independent routes but in the long run there is an equilibrium relationship.
The definition of integration is given below:

"Two or more non-stationary time series are cointegrated if there is a linear combination between these time series that is stationary."

There are two basic methods that are used to test the cointegration of two or more variables. The method that we can use to investigate the existence of a long-term causal relationship between two variables is the Granger test (C. W. J. Granger, 1969). In order for this test to be implemented, a requirement is the existence of stationary time series. A time series is characterized as stationary, when the mean, variance and autocorrelation (in various time lags) are stable over time, regardless of the point in time at which they are measured, ie they are independent to time. In contrast, for non-stationary time series, the parameters vary over time. It is important for a time series to be stationary, as in the case of non-stationarity, the conclusions cannot be generalized beyond the time period under observation and therefore, we are not able to predict the behavior of the time series in the long run.
When financial time series are non-stationary, researchers are investigating methods to turn them into stationary. With the cointegration method we can better approach the existence of a long-term equilibrium relationship between our variables, without the risk of losing useful information. The concept of integration was first proposed by Granger (C. Granger, 1981). He said that if two chronological series Xt and Yt are integrated of order d, that is I (d) and there is a linear combination between them that gives us an integrated series lower than the original, let I (db) for b> 0, then according to Engle and Granger (Engle & Granger, 1987), series are integrated in order (db).
According to Engle and Granger (Engle & Granger, 1987), the linear combination of two variables Xt and Yt, I (1) results in a variable I (0), whereby the two variables are cointegrated. By regressing two chronologically non-stationary series, if the residues of the equation (u_t) equal to Yt - λXt, are I (0), then the variables are cointegrated.

Various techniques for determining the short-run relationships between variables, as well as their long-term variability, were applied by Granger. The basis of this method is the idea that the linear combination of two or more non-stationary time series can be a stationary time series. The justification for this analysis comes from economics, which finds that if there is a long-run equilibrium between two variables, for example disposable income and consumption, then their short-term behavior may differ from the long-term, but will progressively change accordingly to the long-term balance. This concept was named by Granger as a concept of cointegration.

Granger, with the help of Engle, published a groundbreaking article in 1987, not only on Economics but also on the Econometric approach to quantitative investigation. Its main topic was statistical techniques for controlling integration, as well as the method of estimating linear systems which includes the concept of long-term change. In particular, estimating linear relationships between two non-stationary variables requires two steps. First their long-term change is pointed out and it is examined whether the concept of integration is valid and then the so-called ECM error correction model is calculated, which regulates the short-term behavior of the variables adapted from their long-term change. The concept of cointegration was later extended by Johansen to investigate the cointegration of two or more variables. Johansen's methodology uses VAR vector autoregression models.

4.2.1 Johansen test

Johansen test (1988) is a more enriched method than the originally developed by Granger that allows more than one cointegrating relationships. In general, if there are q variables I (1), then there is a q-1 of integrated vectors. Johansen's approach makes it possible to find the maximum number of integrated vectors that exist between a group of variables.

The VAR model (or vector autoregression model), that the Johansen method is based on, is a system of equations where each variable is affected not only by its previous values but also by the previous values of the other variables in the system. Johansen's approach allows us to test several equations at the same time, something that is not happening with the Engle and Granger methods, where one equation must be tested at a time. This is the main reason why the Johansen method has dominated in testing the cointegration of many variables (Johansen, 1988).

Assuming that we have the following VAR model with n variables:

$$Z_t = c + \sum_{j=1}^{p} A_j Z_{t-j} + e_t \quad (4.3)$$

In order to proceed to the estimation of the above VAR model we consider the following hypotheses:

a) the n variables included in the vector of the endogenous variables Zt are integrated of order one Zt ~ I (1) or zero order Zt ~ I (0)

b) c is the (n x 1) vector of fixed terms

c) with A_j are the (n x n) matrices of the coefficients of endogenous variables with time lag

d) e_t is the (n x 1) vector of residues for which we assume that:

E (et) = 0 and COV (et, es) = E (etes΄) = Σδts (where δts is the Kronecker delta)

This method was named after Soren Johansen and involves, using the method of maximum probability, the estimation of a VAR model. Then the integration test is performed by performing the trace test and the maximum value test (λ-max test). The maximum eigenvalue equation is the following:

$$\lambda_{max} = (r, r+1) = -nlog(1 - \hat{\lambda}_{r+1}) \qquad (4.4)$$

H_0: rank ≤ r, H_1: rank=r+1

This test estimates the eigenvalues ($\hat{\lambda}_{r+1}$). When the (r+1)th estimated eigenvalue is accepted to be zero, then smaller eigenvalues are also accepted to be zero.

The trace equation is the following:

$$\lambda_{trace}(r) = -n\sum_{j=r+1}^{m} \log\,(1 - \hat{\lambda}_j) \qquad (4.5)$$

Where only the smallest (m-r) estimated eigenvalues ($\lambda = \lambda_r$+1, …, λ_m) are significantly different from zero. The Ho for the trace test is that there are (m − r) cointegrating vectors.

4.3 Error Correction Vector Autoregression (EC VAR)

Confirmation of the presence of cointegration and the direction is concluded through the value of the error correction term which must be negative and statistically significant. Moreover, the value of this term % shows us the speed of adjustment of equilibrium over a period of time.

As discussed in the previous section, when two variables have a long-term relationship, they are also cointegrated. However, these variables in the short run could present an imbalance relationship, where the residues represent the imbalance errors - balancing errors. This short-term relationship is formulated through a dynamic correction model Error Correction Model (ECM), which links the long-run to the short-run relationship of the variables. ECM are useful to estimate short-term and long-term effects of one time series on another. Moreover, ECMs estimate the speed that a dependent variable needs to return to equilibrium after a change in another variable.

The VAR model (4.5) can be transformed into the form

$$\Delta Z_t = c + \sum_{j-1}^{p-1} Q_j \Delta Z_{t-j} + \Pi Z_{t-p} + e_t \qquad (4.6)$$

$$\text{where } Q_j = \sum_{k=1}^{j} A_k - I \ \text{ and } \Pi = \sum_{k=1}^{p} A_k - I \qquad (4.7)$$

The above transformation is also known as cointegrating transformation while the model is called VECM (Vector Error Correction Model) and is the general form of a multivariate error correction model.

4.4 Granger causality

In economics, the causality test can be used to detect the ability to use prior values of a time series in order to predict the future values of another time series. This connection of the variables is totally different for the correlation of two variables. A correlation between variables in a dataset, could be causal or not and it usually refers to a linear relation of two variables. Causal relationship means that a variation in a variable is the result of a change in another variable.

The detection of a causal relationship in variables at economic studies is a very important finding, as it can be used to predict future values at forecasting models. The application of a simple regression on two economic variables can show us the existence of dependence between them, in the sense that the independent determines the dependent. A statistically significant factor, however, does not mean that there is a causal relationship between the variables. We need to find the direction of the dependence, that is, which changes lead and are reflected on the variable which is caused by the first.

The complexity behind this process led Granger (1969) to develop his theory by incorporating the concept of causality into economics as the well-known "Granger

Causality". It is defined as the phenomenon where previous information about the values of one variable (Xt) helps to better predict the values of another variable (Yt). Then we say that Xt justifies the variable Yt according to Granger. According to Granger's definition, if we include in Yt's forecast only its previous values and ignore the previous values of Xt the prediction will turn out to be less accurate than if we included the values of Xt in the analysis.

In an article, Granger (1988) gave a vector explanation of his theory by making the hypothesis of three vectors of xt, yt and wt.

In recent years, the most common method of testing the causal relationship between two variables is that proposed by Granger, a statistical test also known as the Granger Causality test. We consider two stationary time series, Y and X, with a simple causal model (C. W. J. Granger, 1969):

$$X_t = a + \sum_{i=1}^{p} \beta_i X_{t-i} + \sum_{i=1}^{p} \gamma_i Y_{t-i} + u_t \qquad (4.8)$$

$$Y_t = a + \sum_{i=1}^{p} b_i Y_{t-i} + \sum_{i=1}^{p} c_i X_{t-i} + v_t \qquad (4.9)$$

The value of Y at time t, is a function of past values but also those of X while in the second equation, the value of X at time t is a function of its previous values and those of X. This is a vector autoregressive model. Using Statistics F we will calculate in both relations the statistical significance of the coefficients bi and ci in order to determine the existence or non-existence of a relationship between the variables. We will compare in each case the value F with that of Fcritical.

To examine the presence of causality from Y to X we omit the terms of Yt from the first equation, and then we have the following equation:

$$X_t = a^* + \sum_{i=1}^{p} \beta_i^* X_{t-i} + u_t^* \quad (4.10)$$

If the addition of the terms Yt reduces the variance of the residues in equation (3.1.1), ie var (ut) ‹var (u * t), then we say that Y affects X causally according to Granger.

It is important to note that the basic patterns of Granger causality test are four. The first is to have an one-way causality from x to y. The second is to have an one-way causality from y to x. The third is a bidirectional (two-way) causality of x and y, which means that we have at the same time causality from x to y and from y to x. The fourth pattern it to have complete lack of causal effects.

4.5 Impulse response and Variance decomposition

At econometrics, impulse response functions describe how the economy reacts over time to external events called shocks, which are formed in the context of a vector autoregression model. The incentives are treated as exogenous from a macroeconomic point of view and include various events such as changes in government spending, fiscal or monetary policy or even changes in productivity. Impulse response functions describe the response of the endogenous macroeconomic variable such as production, consumption, investment and employment at the time of the shock and at subsequent times.

As for the variance decomposition, in econometrics, when we have multivariate time series, the variance decomposition is used to help for a better interpretation fo the VAR model that has been created. Variance decomposition describes the contribution of each variable to the other variables of the autoregression. It defines in what extent the forecast error variance of every variable can be explained when exogenous shocks occur to the other variables.

These techniques use a shock of one standard deviation and identifies how this shock affects the variables of the system. An unpredictable shock in a variable directly affects not only itself, but is transmitted to other endogenous variables of the system.

5. Empirical results and Discussion

5.1 Data

This study uses monthly time-series data to empirically test the long-run and short-run interaction between cotton A index prices (CAI), crude oil price (COIL) and world cotton stocks to world cotton use ratio (STU). The analysis covers the period from October 1990 to February 2021, making a time-series data of 365 observations. The source of the cotton A index prices and crude oil price is the datasets listed in The World Bank Data Catalog cited as "Pink Sheet" Data and more specifically in the section of Commodity monthly prices, as an "xls" file. The source of world cotton stocks and world cotton use is the "World Agricultural Supply and Demand Estimates", a monthly report that is published from the United States Department of Agriculture (USDA). The three variables were used in logarithmic form denoted with an "L" in front of the initial variable name.

Table 2. Summary of the variables that were used

Variable	Symbol	Source
Cotton A index ($/kg)	CAI	World Bank
Crude oil, average ($/bbl)	COIL	World Bank
$\frac{World\ cotton\ stocks}{World\ cotton\ use}$	STU	USDA

5.2 Empirical results

5.2.1 Unit root tests

In order to investigate the integration properties of the three variables, the Dickey-Fuller GLS test and the Kwiatkowski-Phillips-Schmidt-Shin test were used complementary. These two tests were applied firstly on the levels of the variables with constant term and linear trend and then using the first differences of the variables with constant term and linear trend. The unit root tests were conducted before causality tests so as to identify the stationarity properties of the variables. The results of the unit root tests indicated a unit root for the three variables (LCAI, LCOIL and LSTU) in levels at the 1% level of significance and stationarity

in the first differences at the 1% level of significance (Table 3). The results of the unit-root tests allow us to suppose that our variables are integrated of order one or I(1).
Our results are in accordance with similar research, as Cotton A index has been found to be non-stationary by many researchers, as well as world cotton stocks and stocks to use ratio (MacDonald, 2006).

Table 3. Unit root tests results (with constant term and linear trend)

Variables		DF GLS Test	KPSS Test
LCAI	Level	-3.235369	0.181355
	1st Difference	-10.74351	0.028288
LCOIL	Level	-2.185814	0.322534
	1st Difference	-11.68310	0.083552
LSTU	Level	-2.053468	0.182552
	1st Difference	-16.55914	0.049351

Note: The critical values for DF GLS Test for 1%, 5% and 10% are -3.476300, -2.897400 and -2.582950 respectively. The critical values for KPSS Test for 1%, 5% and 10% are 0.216, 0.146 and 0.119 respectively

5.2.2 Structural break test

Since we tried to identify if there are causal links between our variables in the long run, we proceeded to a preliminary analysis using OLS and we tested the residuals of the equation for possible structural breaks as to include them later using the appropriate methodology.
Structural breaks are unexpected changes in the time series that could lead to forecasting errors. The use of this test could identify events on specific dates that something broke the statistical properties of our variables. Especially in studies with time-series data similar to our data (commodity prices, world production, etc.) it is common to use structural break tests. In fact, these dates are matching with significant groundbreaking economic events.
We conducted a stability test (multiple breakpoint test) to identify if any structural break, using the method "Sequential L +1 breaks vs L" with 2 maximum breaks. The method of Bai and Perron (Bai & Perron, 2003) was used in this study to identify break-dates in the variables. We identified two dates of structural breaks: **2009m12** and **2001m03** (Table 4).

Table 4. Results of Multiple breakpoint tests based on Bai & Perron (2003)

Bai-Perron tests of L+1 vs. L sequentially determined breaks

Sequential F-statistic determined breaks: 2

Break Test	F-statistic	Scaled F-statistic	Critical Value**
0 vs. 1 *	108.7558	326.2674	13.98
1 vs. 2 *	54.94303	164.8291	15.72

Break dates:	Sequential	Repartition
1	**2009M12**	2001M03
2	**2001M03**	2010M02

Sample: 1990M10 till 2021M02 (365 observations); Breaking variables: LCOIL, LSTU, C; Break test options: Trimming 0.15, Max. breaks 2, Sig. level 0.05; * Significant at the 0.05 level.

Regarding our variables, we expected that the 2000s commodities boom would be a structural break in our time-series data (Figure 8). Indeed, this is a structural break identified by many researchers as at this period USA exports have significantly surpassed the use of cotton in the states for the first time after about 60 years that about 60% of U.S. cotton was consumed domestically (Isengildina-Massa & MacDonald, 2009). More specifically, this structural change it may be caused by a combination of factors, including the increased export orientation of the U.S. cotton industry following the U.S. textile industry's contraction (Isengildina-Massa & MacDonald, 2009). At the same period (2001) China joined the WTO for a textile trade liberalization.

As for the second identified date of structural break (Figure 8), is attributed to a very important external shock of the international cotton market that has been conducted at 2010, when China's government decided to start a program to increase national stocks of cotton, while at the same time oil prices rise (OECD et al., 2020).

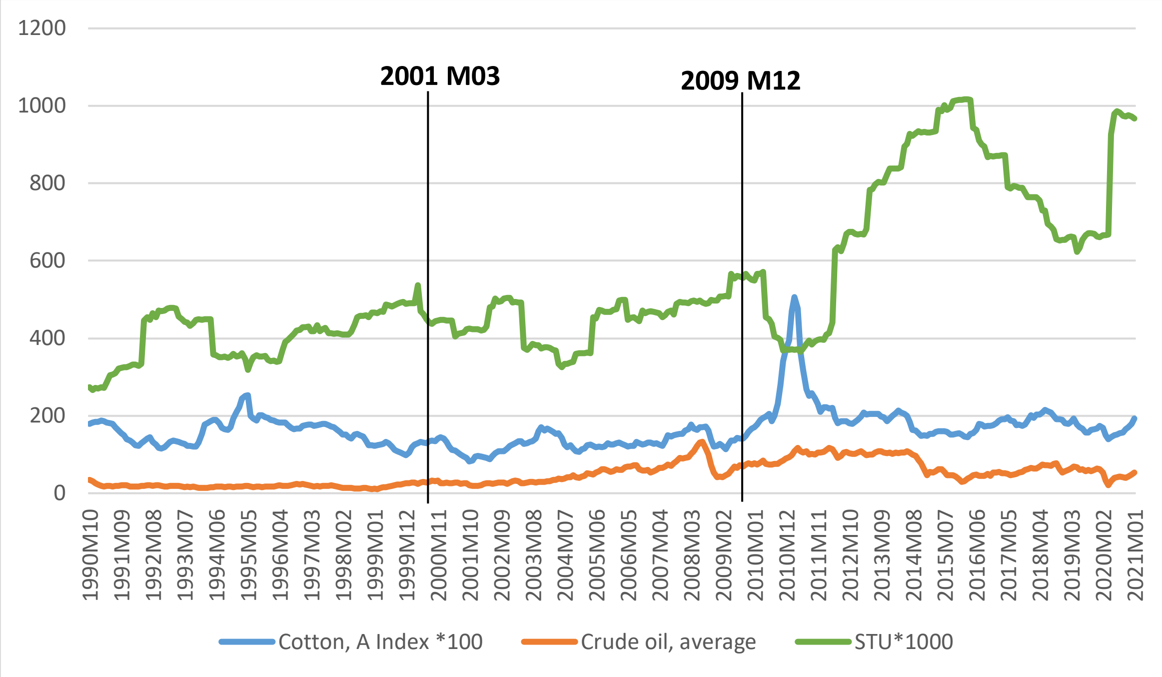

Figure 8. Graphical presentation of the two structural breaks identified from our analysis in comparison to our data

After the identification of structural breaks, we created 4 dummy variables, an impulse and a stability variable for each date. Then we proceeded to the creation of the equation of our 3 variables along with the 4 dummy variables using the method of LS - Least Squares (NLS and ARMA). Further preliminary analysis of the significance of the new created dummy variables, indicated that the two stability variables were statistically significant, while the two impulse variables were non-significant. After this preliminary analysis, the two stability dummy variables **S_2001M03** and **S_2009M12** were incorporated as deterministic variables.

5.2.3 Cointegration test

When our main goal is to investigate if there are any dynamic links between the variables, and when we have found that they are non-stationary in levels, but they are stationary at first differences, there is probably a stable long-run equilibrium relationship that can be detected through Cointegration Test process.

After creating the equation, we proceed to check if this relationship is stable. In order to check this, there are many available methods based on singe equation models or more than one cointegrating relationships. In our research we decided to use the Johansen's cointegration methodology. This method is more appropriate for multivariate analysis. It has been proved that this methodology, which is based on a system of equations is the most consistent. Moreover, since we have many observations (365), there is no problem of lack of reliability of the results due to small number of degrees of freedom.

At the first step of the analysis, we checked various possible cases of combinations regarding the long-term relationship, the cointegrating equation (CE) and the error correction model, in order to comparatively settle on that form which provides some acceptable results. With the help of the selection criteria of the appropriate model, such as Akaike Information Criteria and Schwarz Criteria, we selected the second combination (Intercept with No trend). This choice is based on the Akaike criterion, which indicates that the most appropriate specialization is the one that includes intercept but no trend. Among the 5 possible combinations that can be evaluated of data trend (none, linear, quadratic), test type (no intercept, intercept, no trend, trend), the criterion of Akaki indicated that the best choice was Intercept with No trend. This combination had 1 stable relationship in Trace and Max-Eigenvalue.

Table 5. Results of Cointegration test

Unrestricted Cointegration Rank Test (**Trace**)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.079177	47.96070	35.19275	**0.0013**
At most 1	0.029322	18.26529	20.26184	0.0920
At most 2	0.020758	7.551472	9.164546	0.1002
Trace test indicates 1 cointegrating eqn(s) at the 0.05 level				
Unrestricted Cointegration Rank Test (**Maximum Eigenvalue**)				
Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.079177	29.69541	22.29962	**0.0039**
At most 1	0.029322	10.71382	15.89210	0.2737
At most 2	0.020758	7.551472	9.164546	0.1002
Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level				

Sample: 1991M03 - 2021M02 (360 after adjustments); Series: LCAI LCOIL LSTU; Exogenous series: S_2001M03 S_2009M12; Lags interval (in first differences): 1 to 4.
*: rejection of the hypothesis at the 0.05 level; **: p-values (MacKinnon et al., 1999)

In the selected model (Intercept with No trend) we apply the cointegration test which is done with the help of Trace and Max-Eigenvalue test. The Trace test (Table 5) indicates that we reject Ho (0.0013) that we have none cointegrating vector and that we accept Ho (p=0.0920) that we have at most 1 cointegrating vector. We conclude that we have 1 cointegrating vector. This conclusion is in line with the results of the Maximum Eigenvalue test which rejects Ho (p=0.0039) that we have none cointegrating vector and accepts Ho (p=0.1575) that we have at most 1 cointegrating vector. The results of this analysis indicate that there is at most 1 cointegrating equation, which means a long-term relationship of the variables. The long-run coefficients for the one detected cointegrating relationship is presented in Table 6. Johansen cointegration methodology has been applied in a similar research that studied the dynamic interactions of ethanol, crude oil and corn price and resulted in a system of estimation approaches to cointegration (Dimitriadis & Katrakilidis, 2020).

Table 6. Long-run coefficients

Dependent variable	Independent variables		Constant
LCAI	LCOIL	LSTU	
	0.192432	-0.447347	0.536161
	(-2.68566)	(3.80294)	(2.45973)

Note: t statistics are presented in parenthesis.

The next step is to evaluate the estimated error correction terms. We conducted this analysis in order to obtain information regarding the direction of causality. Our results indicate that we have a negative and significant error correction term for the variables LCAI and LSTU which confirms the existence of long-run causal effects. So, there is a long-run causal effect that is directed towards the LCAI and another long-run causal effect that is directed towards the LSTU. LCAI is affected by the other two variables in the long-run and LSTU by the other two variables in the long-run. LCOIL, however, is not affected in the long run by the other two and it is considered as exogenous. In addition, the value of this term % shows us the speed of adjustment of equilibrium over a period of time.

Table 7. Vector Error Correction Estimates

Error Correction:	D(LCAI)	D(LCOIL)	D(LSTU)
CointEq1	-0.079204	0.013777	-0.045542
	(0.01818)	(0.03348)	(0.01832)
	[-4.35548]	[0.41145]	[-2.48536]
D(LCAI(-1))	0.563114	0.288268	0.021071
	(0.05339)	(0.09830)	(0.05379)
	[10.5479]	[2.93254]	[0.39170]
D(LCAI(-2))	-0.052787	-0.130090	-0.032968
	(0.06131)	(0.11290)	(0.06178)
	[-0.86093]	[-1.15229]	[-0.53361]
D(LCAI(-3))	0.040053	0.033882	-0.060650
	(0.06071)	(0.11179)	(0.06118)
	[0.65971]	[0.30308]	[-0.99138]
D(LCAI(-4))	0.056171	0.183664	0.067547
	(0.05571)	(0.10258)	(0.05614)
	[1.00823]	[1.79039]	[1.20323]
D(LCOIL(-1))	0.005653	0.260999	-0.031306
	(0.02939)	(0.05411)	(0.02961)
	[0.19237]	[4.82374]	[-1.05727]
D(LCOIL(-2))	-0.005061	-0.080910	-0.013452
	(0.03037)	(0.05592)	(0.03060)
	[-0.16666]	[-1.44697]	[-0.43959]
D(LCOIL(-3))	-0.004604	-0.063706	-0.017532
	(0.03032)	(0.05584)	(0.03056)
	[-0.15182]	[-1.14096]	[-0.57379]
D(LCOIL(-4))	-0.038891	-0.078672	-0.013780
	(0.02908)	(0.05354)	(0.02930)
	[-1.33758]	[-1.46950]	[-0.47035]
D(LSTU(-1))	0.015423	0.108777	0.063210
	(0.05307)	(0.09771)	(0.05347)
	[0.29063]	[1.11324]	[1.18211]

D(LSTU(-2))	0.069366	0.015200	0.019732
	(0.05304)	(0.09766)	(0.05344)
	[1.30787]	[0.15565]	[0.36922]
D(LSTU(-3))	-0.106077	0.043593	0.040696
	(0.05284)	(0.09730)	(0.05325)
	[-2.00745]	[0.44804]	[0.76430]
D(LSTU(-4))	-0.086652	0.072275	0.034388
	(0.05298)	(0.09755)	(0.05338)
	[-1.63562]	[0.74092]	[0.64418]
S_2001M03	-0.023302	0.012752	-0.012100
	(0.00734)	(0.01352)	(0.00740)
	[-3.17409]	[0.94340]	[-1.63577]
S_2009M12	0.040920	-0.019141	0.024840
	(0.01128)	(0.02077)	(0.01137)
	[3.62741]	[-0.92152]	[2.18526]
R-squared	0.313755	0.138882	0.050964
Adj. R-squared	0.285907	0.103938	0.012453
Sum sq. resids	0.744567	2.524366	0.755996
S.E. equation	0.046456	0.085539	0.046811
F-statistic	11.26683	3.974429	1.323348
Log likelihood	601.7723	382.0027	599.0302
Akaike AIC	-3.259846	-2.038904	-3.244612
Schwarz SC	-3.097925	-1.876983	-3.082691
Mean dependent	0.000247	0.003345	0.003506
S.D. dependent	0.054975	0.090364	0.047105

Sample: 1991M03 2021M02 (360 observations after adjustments); Standard errors in ( ) & t-statistics in [ ];

5.2.4 Granger Causality test

After the creation of the equation with the long-run relationship, we conducted a Granger Causality test. The results of this test indicate that there is causality only in the case of LCAI that affects LCOIL statistically significantly (p=0.0162). That means that we have a short-run causal effect from LCAI to LCOIL. The volatility of oil price has been found to have a significant effect on cotton price. Especially cash crops, such as cotton, the impact of oil price shocks is stronger than the shock at food crops and the explanation for this behavior is the fact that the cash crops are more dependent on foreign imports than food crops such as wheat and maize (Zhang & Qu, 2015). Oil price fluctuations play a crucial role for the world cotton market, as oil price shocks affect the cost of production inputs, the cost of transportation, as well as the processing process (Zhang & Qu, 2015).

Table 8. Granger Causality test

Dependent variable: D(LCAI)			
Excluded	Chi-sq	df	Prob.
D(LCOIL)	2.206243	4	0.6979
D(LSTU)	8.575800	4	0.0726
All	11.07032	8	0.1977
Dependent variable: D(LCOIL)			
Excluded	Chi-sq	df	Prob.
D(LCAI)	12.15972	4	0.0162
D(LSTU)	2.225772	4	0.6943
All	13.28862	8	0.1023
Dependent variable: D(LSTU)			
Excluded	Chi-sq	df	Prob.
D(LCAI)	2.757415	4	0.5992
D(LCOIL)	2.515784	4	0.6418
All	6.049313	8	0.6417

Sample: 1990M10 2021M02 (360 observations).

5.2.5 Variance Decompositions

We used variance decompositions in order to detect for each variable, which is the variable that explains the largest percentage of variance. This is also a technique that is indirectly a causality test like the Granger Causality test. So, we would like to confirm the results that we found from the cointegration analysis. This analysis presents how this causality is formed during a period of time, which in our research is 24 months.

Table 9. Variance decomposition of LCAI, LCOIL and LSTU

LCAI				
Period	S.E.	LCAI	LCOIL	LSTU
1	0.046456	100.0000	0.000000	0.000000
6	0.167698	97.55060	0.268638	2.180758
12	0.216619	88.64393	1.510022	9.846049
18	0.238088	81.24370	3.808204	14.94809
24	0.253779	75.80480	5.942109	18.25309
LCOIL				
Period	S.E.	LCAI	LCOIL	LSTU
1	0.085539	4.089026	95.91097	0.000000
6	0.250037	13.84669	85.37677	0.776537
12	0.344537	15.58847	83.82879	0.582740
18	0.416946	13.56581	85.99557	0.438621
24	0.478218	11.98923	87.62631	0.384458
LSTU				
Period	S.E.	LCAI	LCOIL	LSTU
1	0.046811	0.889011	0.346705	98.76428
6	0.125623	9.392259	0.096381	90.51136
12	0.191950	18.98643	0.070441	80.94313
18	0.246227	23.62066	0.087552	76.29179
24	0.291587	25.82570	0.097541	74.07676

The most important variable that we want to explain is LCAI. Our results show that in fact LSTU explains LCAI and this explanation reaches only up to 20% (Table 9). Moreover, the 1st year does not show any significant explanatory ability of the behavior, while in the 2nd year there is a moderate possibility that the rate of explanation of the behavior (of the variability of LCAI) starts from about 13% and reaches 23% at the end of the second year. As for LCOIL, we already knew (from the estimated error correction terms) that it is not affected by any of the other two variables. It starts from 95% at the 1st year and drops to 87 at the end of the 2nd year, which means that it is an exogenous variable in the system. For the other two variables the percentage of explanatory is minor. Concerning LSTU, it is mainly explained by LCAI.

In a recent study, that used daily data from 2004 to 2014, it was found that the rise and fall in the oil price has the same effects on the price of cotton (Zhang & Qu, 2015).

5.2.6 Impulse response

This analysis presents the effect of a shock and when this shock peaks, at every other variable. We chose to investigate the response of the variables to the shock at a horizon of 24 months. We can see that a shock at the value of stocks to use ratio, is starting to reduce the price of cotton approximately 4 months after the event, maximises about 6 months later and then remains the same for the examined 24 month period (Figure 9). A shock at the price of crude oil leads to a mild increase of the cotton price during the 24 month period. Another important finding is how the price of cotton affects the stocks to use ratio, which is a ratio based on the fundamentals of the cotton market. We can see that a shock at the price of cotton, results in an immediate decrease at the value of stocks to use ratio, which takes its maximum value about 14 months after the shock.

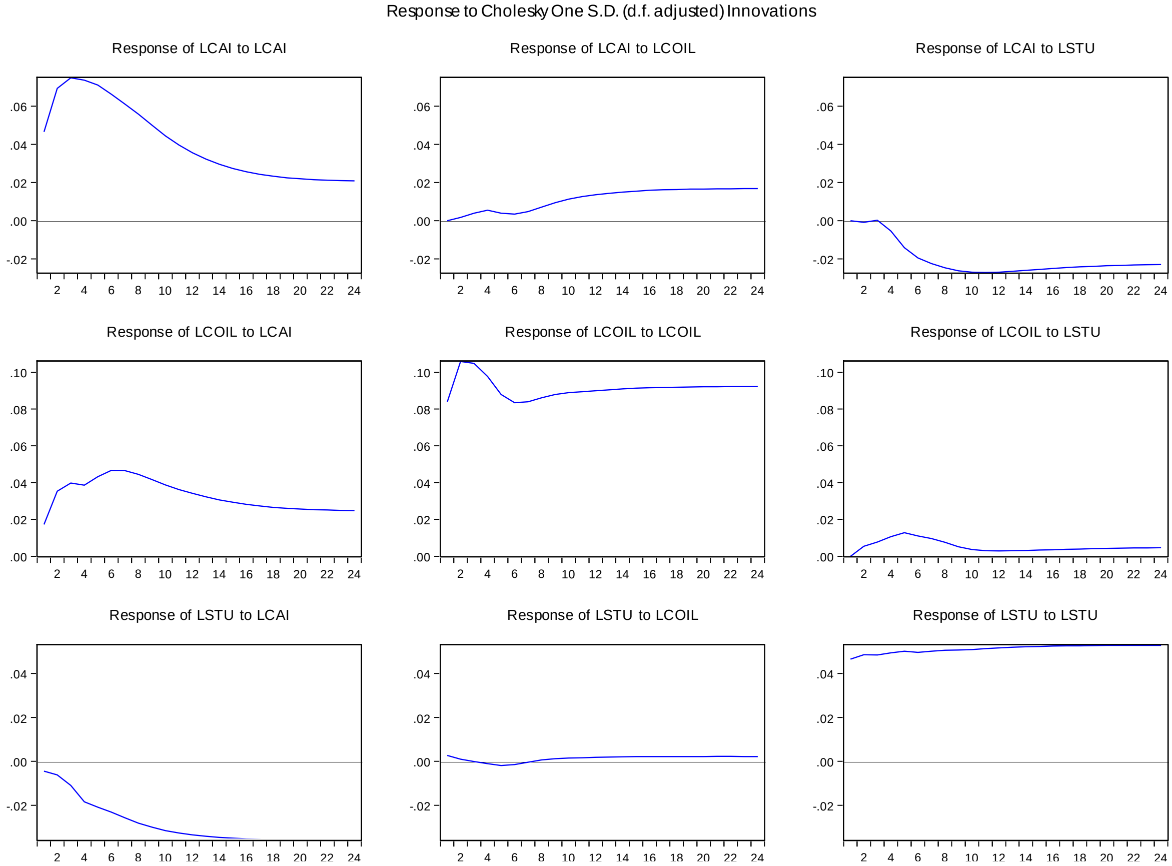

Figure 9. Impulse response of LCAI, LCOIL and LSTU at a horizon of 24 months

6. Conclusions

At this study, the dynamic relationships among cotton price, crude oil price and world cotton stocks to world cotton use ratio in the long-run and in the short-run were investigated. The empirical analysis was based on the Johansen cointegration methodology. The results of the analysis revealed the existence of significant long-run causal effect that is directed towards the cotton prices and another long-run causal effect that is directed towards the stocks to use ratio. Cotton price is affected by the other two variables in the long-run and stocks to use ratio by the other two variables in the long-run. Crude oil price, however, is not affected in the long run by the other two and it is considered as exogenous to the system. Moreover, the Granger causality test indicated that there is causality only in the case of cotton prices that affects crude oil price significantly, which means that we have a short-run causal effect from cotton price to crude oil price. Finally, impulse response and variance decomposition analysis were applied to our results to study the out of sample forecasting behavior of the system. The results indicated that after a shock the value of stocks to use ratio, is starting to reduce the price of cotton approximately 4 months after the event and maximizes about 6 months later, while a shock at the price of crude oil leads to a mild increase of the cotton price during the 24 month period. As for the explanatory power of the variables, cotton stocks to use ratio explains only up to 20% of the behavior of cotton prices.

It is important to note that there is little research published regarding the dynamic interactions among cotton price and variables such as crude oil price and stocks to use ratio. To the best of our knowledge, there is no published work that uses the same methodology that we use in this study. The use of Johansen cointegration test, which is a methodology widely accepted for its reliable results contributes to the econometrics science with novel results, regarding the ability to make models that can forecast the movements of cotton price in the long-run and in the short-run.

Bibliography

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*(1), 1–22. https://doi.org/10.1002/jae.659

Chen, D. T., & Bessler, D. A. (1990). Forecasting monthly cotton price: Structural and time series approaches. *International Journal of Forecasting*, *6*(1), 103–113. https://doi.org/10.1016/0169-2070(90)90101-G

Coakley, J., Dollery, J., & Kellard, N. (2011). Long memory and structural breaks in commodity futures markets. *Journal of Futures Markets*, *31*(11), 1076–1113. https://doi.org/10.1002/fut.20502

Dimitriadis, D., & Katrakilidis, C. (2020). An empirical analysis of the dynamic interactions among ethanol, crude oil and corn prices in the US market. *Annals of Operations Research*, *294*(1), 47–57. https://doi.org/10.1007/s10479-018-2832-0

Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica*, *64*(4), 813–836. JSTOR. https://doi.org/10.2307/2171846

Engle, R. F., & Granger, C. W. J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, *55*(2), 251–276. JSTOR. https://doi.org/10.2307/1913236

Ghorbel, A., Hamma, W., & Jarboui, A. (2017). Dependence between oil and commodities markets using time-varying Archimedean copulas and effectiveness of hedging strategies. *Journal of Applied Statistics*, *44*(9), 1509–1542. https://doi.org/10.1080/02664763.2016.1155107

Granger, C. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, *16*(1), 121–130.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424–438. JSTOR. https://doi.org/10.2307/1912791

Husain, M., & Bowman, C. (2004). *Forecasting commodity prices: Futures versus judgment. Number 4-41*. International Monetary Fund Working Paper.

ICAC. (2016). *COST OF PRODUCTION OF RAW COTTON*. https://www.icac.org/Content/PublicationsPdf%20Files/9697a6b2_6d66_4111_b66d_2d2d0efea4b8/cost-of-production2016.pdf.pdf

Isengildina-Massa, O., & MacDonald, S. (2009). U.S. Cotton Prices and the World Cotton Market; Forecasting and Structural Change. In *Economic Research Report* (No. 55950; Economic Research Report). United States Department of Agriculture, Economic Research Service. https://ideas.repec.org/p/ags/uersrr/55950.html

Katrakilidis, C., Mattas, K., & Polymeros, K. (2005). *Structural Changes and Trend Behaviour of the European Cotton*. *2 (2)*, 207–215.

Khan, M. A., Wahid, A., Ahmad, M., Tahir, M. T., Ahmed, M., Ahmad, S., & Hasanuzzaman, M. (2020). World Cotton Production and Consumption: An Overview. In S. Ahmad & M. Hasanuzzaman (Eds.), *Cotton Production and Uses: Agronomy, Crop Protection, and Postharvest Technologies* (pp. 1–7). Springer Singapore. https://doi.org/10.1007/978-981-15-1472-2_1

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1), 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y

MacDonald, S. (2006). *Cotton Price Forecasting and Structural Change*.

MacDonald, S., & Leslie, M. (2018). *Long Run Trends and Fluctuations In Cotton Prices*.

MacKinnon, J., Haug, A., & Michelis, L. (1999). Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration. *Journal of Applied Econometrics*, *14*(5), 563–577.

OECD, Food, & Nations, A. O. of the U. (2020). *OECD-FAO Agricultural Outlook 2020-2029*. https://doi.org/10.1787/1112c23b-en

Oladosu, G. A., Kline, K. L., & Langeveld, J. W. A. (2021). Structural Break and Causal Analyses of U.S. Corn Use for Ethanol and Other Corn Market Variables. *Agriculture*, *11*(3), 267. https://doi.org/10.3390/agriculture11030267

Olen, C., & Andersson, T. N. (2013). *Can the forecast of the cotton price be improved using a model based upon economic variables?* Lund University.

Power, G. J., & Robinson, J. R. C. (2009). Cotton Futures Dynamics: Structural Change, Index Traders and the Returns to Storage. *Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. St. Louis, MO.*

Zhang, C., & Qu, X. (2015). The effect of global oil price shocks on China's agricultural commodities. *Energy Economics*, *51*, 354–364. https://doi.org/10.1016/j.eneco.2015.07.012

Appendix A: Tables with detailed results of the statistical analysis

1) Unit root tests at level and 1st diference for LCAI

Dickey-Fuller GLS (ERS) test at Level for LCAI

Null Hypothesis: LCAI has a unit root				
Exogenous: Constant, Linear Trend				
Lag Length: 1 (Automatic - based on SIC, maxlag=8)				
				t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic				-3.235369
Test critical values:	1% level			-3.476300
	5% level			-2.897400
	10% level			-2.582950
*(Elliott et al., 1996) (Table 1)				
DF-GLS Test Equation on GLS Detrended Residuals				
Dependent Variable: D(GLSRESID)				
Method: Least Squares				
Date: 04/04/21 Time: 17:48				
Sample (adjusted): 1990M12 2021M02				
Included observations: 363 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
GLSRESID(-1)	-0.028757	0.008888	-3.235369	0.0013
D(GLSRESID(-1))	0.513999	0.045276	11.35265	0.0000
R-squared	0.269714	Mean dependent var		0.000402
Adjusted R-squared	0.267691	S.D. dependent var		0.054756
S.E. of regression	0.046857	Akaike info criterion		-3.277919
Sum squared resid	0.792619	Schwarz criterion		-3.256462
Log likelihood	596.9423	Hannan-Quinn criter.		-3.269390
Durbin-Watson stat	1.920055			

Dickey-Fuller GLS (ERS) test at 1st difference for LCAI

Null Hypothesis: D(LCAI) has a unit root	
Exogenous: Constant, Linear Trend	
Lag Length: 0 (Automatic - based on SIC, maxlag=8)	
	t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic	-10.74351

Test critical values:	1% level			-3.476300
	5% level			-2.897400
	10% level			-2.582950

*(Elliott et al., 1996) (Table 1)

DF-GLS Test Equation on GLS Detrended Residuals
Dependent Variable: D(GLSRESID)
Method: Least Squares
Date: 04/04/21 Time: 17:50
Sample (adjusted): 1990M12 2021M02
Included observations: 363 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GLSRESID(-1)	-0.484770	0.045122	-10.74351	0.0000

R-squared	0.241758	Mean dependent var	0.000137
Adjusted R-squared	0.241758	S.D. dependent var	0.054790
S.E. of regression	0.047710	Akaike info criterion	-3.244619
Sum squared resid	0.823985	Schwarz criterion	-3.233891
Log likelihood	589.8983	Hannan-Quinn criter.	-3.240354
Durbin-Watson stat	1.905502		

Kwiatkowski-Phillips-Schmidt-Shin at Level for LCAI

Null Hypothesis: LCAI is stationary
Exogenous: Constant, Linear Trend
Bandwidth: 15 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.181355
Asymptotic critical values*:	1% level	0.216000
	5% level	0.146000
	10% level	0.119000

(Kwiatkowski et al., 1992) (Table 1)

Residual variance (no correction)	0.062918
HAC corrected variance (Bartlett kernel)	0.755692

KPSS Test Equation
Dependent Variable: LCAI
Method: Least Squares
Date: 04/04/21 Time: 17:52
Sample: 1990M10 2021M02
Included observations: 365

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.318483	0.026277	12.12029	0.0000
@TREND(«1990M10»)	0.000751	0.000125	6.008557	0.0000

R-squared	0.090460	Mean dependent var	0.455122
Adjusted R-squared	0.087954	S.D. dependent var	0.263374
S.E. of regression	0.251525	Akaike info criterion	0.082914
Sum squared resid	22.96510	Schwarz criterion	0.104283
Log likelihood	-13.13184	Hannan-Quinn criter.	0.091407
F-statistic	36.10275	Durbin-Watson stat	0.047272
Prob(F-statistic)	0.000000		

Kwiatkowski-Phillips-Schmidt-Shin at 1st difference for LCAI

Null Hypothesis: D(LCAI) is stationary
Exogenous: Constant, Linear Trend
Bandwidth: 4 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.028288
Asymptotic critical values*:	1% level	0.216000
	5% level	0.146000
	10% level	0.119000

(Kwiatkowski et al., 1992) (Table 1)

Residual variance (no correction)	0.002979
HAC corrected variance (Bartlett kernel)	0.006045

KPSS Test Equation
Dependent Variable: D(LCAI)
Method: Least Squares
Date: 04/04/21 Time: 17:53
Sample (adjusted): 1990M11 2021M02
Included observations: 364 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.002867	0.005749	-0.498712	0.6183
@TREND(«1990M10»)	1.77E-05	2.73E-05	0.647769	0.5175

R-squared	0.001158	Mean dependent var	0.000360
Adjusted R-squared	-0.001601	S.D. dependent var	0.054686
S.E. of regression	0.054729	Akaike info criterion	-2.967354
Sum squared resid	1.084299	Schwarz criterion	-2.945941

Log likelihood	542.0585	Hannan-Quinn criter.	-2.958844
F-statistic	0.419605	Durbin-Watson stat	1.002223
Prob(F-statistic)	0.517545		

2) Unit root tests at level and 1st diference for LCOIL

Dickey-Fuller GLS (ERS) test at Level for LCOIL

Null Hypothesis: LCOIL has a unit root				
Exogenous: Constant, Linear Trend				
Lag Length: 1 (Automatic - based on SIC, maxlag=8)				
				t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic				-2.185814
Test critical values:	1% level			-3.476300
	5% level			-2.897400
	10% level			-2.582950
*(Elliott et al., 1996) (Table 1)				
DF-GLS Test Equation on GLS Detrended Residuals				
Dependent Variable: D(GLSRESID)				
Method: Least Squares				
Date: 04/04/21 Time: 17:55				
Sample (adjusted): 1990M12 2021M02				
Included observations: 363 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
GLSRESID(-1)	-0.018151	0.008304	-2.185814	0.0295
D(GLSRESID(-1))	0.314056	0.049999	6.281234	0.0000
R-squared	0.104706	Mean dependent var		-0.000203
Adjusted R-squared	0.102226	S.D. dependent var		0.091547
S.E. of regression	0.086742	Akaike info criterion		-2.046270
Sum squared resid	2.716208	Schwarz criterion		-2.024813
Log likelihood	373.3979	Hannan-Quinn criter.		-2.037741
Durbin-Watson stat	1.951419			

Dickey-Fuller GLS (ERS) test at 1st difference for LCOIL

Null Hypothesis: D(LCOIL) has a unit root	
Exogenous: Constant, Linear Trend	
Lag Length: 0 (Automatic - based on SIC, maxlag=8)	
	t-Statistic

Elliott-Rothenberg-Stock DF-GLS test statistic		-11.68310
Test critical values:	1% level	-3.476300
	5% level	-2.897400
	10% level	-2.582950

*(Elliott et al., 1996) (Table 1)

DF-GLS Test Equation on GLS Detrended Residuals
Dependent Variable: D(GLSRESID)
Method: Least Squares
Date: 04/04/21 Time: 17:56
Sample (adjusted): 1990M12 2021M02
Included observations: 363 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GLSRESID(-1)	-0.548035	0.046908	-11.68310	0.0000

R-squared	0.273812	Mean dependent var	0.000187
Adjusted R-squared	0.273812	S.D. dependent var	0.107928
S.E. of regression	0.091972	Akaike info criterion	-1.931908
Sum squared resid	3.062122	Schwarz criterion	-1.921180
Log likelihood	351.6413	Hannan-Quinn criter.	-1.927643
Durbin-Watson stat	2.019750		

Kwiatkowski-Phillips-Schmidt-Shin at Level for LCOIL

Null Hypothesis: LCOIL is stationary
Exogenous: Constant, Linear Trend
Bandwidth: 15 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.322534
Asymptotic critical values*:	1% level	0.216000
	5% level	0.146000
	10% level	0.119000

(Kwiatkowski et al., 1992) (Table 1)

Residual variance (no correction)	0.183602
HAC corrected variance (Bartlett kernel)	2.420894

KPSS Test Equation
Dependent Variable: LCOIL
Method: Least Squares
Date: 04/04/21 Time: 17:56
Sample: 1990M10 2021M02
Included observations: 365

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.770748	0.044887	61.72674	0.0000
@TREND("1990M10")	0.004912	0.000213	23.01399	0.0000

R-squared	0.593343	Mean dependent var	3.664770
Adjusted R-squared	0.592223	S.D. dependent var	0.672853
S.E. of regression	0.429667	Akaike info criterion	1.153851
Sum squared resid	67.01473	Schwarz criterion	1.175220
Log likelihood	-208.5778	Hannan-Quinn criter.	1.162344
F-statistic	529.6439	Durbin-Watson stat	0.045501
Prob(F-statistic)	0.000000		

Kwiatkowski-Phillips-Schmidt-Shin at 1st difference for LCOIL

Null Hypothesis: D(LCOIL) is stationary
Exogenous: Constant, Linear Trend
Bandwidth: 6 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.083552
Asymptotic critical values*:	1% level	0.216000
	5% level	0.146000
	10% level	0.119000

(Kwiatkowski et al., 1992) (Table 1)

Residual variance (no correction)	0.008363
HAC corrected variance (Bartlett kernel)	0.011228

KPSS Test Equation
Dependent Variable: D(LCOIL)
Method: Least Squares
Date: 04/04/21 Time: 17:57
Sample (adjusted): 1990M11 2021M02
Included observations: 364 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.001056	0.009633	-0.109648	0.9127
@TREND("1990M10")	1.42E-05	4.57E-05	0.311166	0.7559

R-squared	0.000267	Mean dependent var	0.001541
Adjusted R-squared	-0.002494	S.D. dependent var	0.091590
S.E. of regression	0.091704	Akaike info criterion	-1.935020

Sum squared resid	3.044291	Schwarz criterion	-1.913607
Log likelihood	354.1736	Hannan-Quinn criter.	-1.926509
F-statistic	0.096824	Durbin-Watson stat	1.385163
Prob(F-statistic)	0.755853		

3) Unit root tests at level and 1st diference for LSTU

Dickey-Fuller GLS (ERS) test at Level for LSTU

Null Hypothesis: LSTU has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=8)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-2.053468
Test critical values:	1% level	-3.476400
	5% level	-2.897200
	10% level	-2.582600

*(Elliott et al., 1996) (Table 1)

DF-GLS Test Equation on GLS Detrended Residuals
Dependent Variable: D(GLSRESID)
Method: Least Squares
Date: 04/04/21 Time: 17:58
Sample (adjusted): 1990M11 2021M02
Included observations: 364 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GLSRESID(-1)	-0.023167	0.011282	-2.053468	0.0407

R-squared	0.011447	Mean dependent var	0.000280
Adjusted R-squared	0.011447	S.D. dependent var	0.046880
S.E. of regression	0.046611	Akaike info criterion	-3.291237
Sum squared resid	0.788633	Schwarz criterion	-3.280531
Log likelihood	600.0052	Hannan-Quinn criter.	-3.286982
Durbin-Watson stat	1.822856		

Dickey-Fuller GLS (ERS) test at 1st difference for LSTU

Null Hypothesis: D(LSTU) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=8)

		t-Statistic

Elliott-Rothenberg-Stock DF-GLS test statistic		-16.55914
Test critical values:	1% level	-3.476300
	5% level	-2.897400
	10% level	-2.582950

*(Elliott et al., 1996) (Table 1)

DF-GLS Test Equation on GLS Detrended Residuals
Dependent Variable: D(GLSRESID)
Method: Least Squares
Date: 04/04/21 Time: 17:59
Sample (adjusted): 1990M12 2021M02
Included observations: 363 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GLSRESID(-1)	-0.862140	0.052064	-16.55914	0.0000

R-squared	0.431001	Mean dependent var	-3.37E-05
Adjusted R-squared	0.431001	S.D. dependent var	0.063751
S.E. of regression	0.048089	Akaike info criterion	-3.228790
Sum squared resid	0.837131	Schwarz criterion	-3.218062
Log likelihood	587.0255	Hannan-Quinn criter.	-3.224526
Durbin-Watson stat	2.019352		

Kwiatkowski-Phillips-Schmidt-Shin test at Level for LSTU

Null Hypothesis: LSTU is stationary
Exogenous: Constant, Linear Trend
Bandwidth: 15 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.182552
Asymptotic critical values*:	1% level	0.216000
	5% level	0.146000
	10% level	0.119000

*(Kwiatkowski et al., 1992) (Table 1)

Residual variance (no correction)	0.041259
HAC corrected variance (Bartlett kernel)	0.532264

KPSS Test Equation
Dependent Variable: LSTU
Method: Least Squares
Date: 04/04/21 Time: 18:00
Sample: 1990M10 2021M02

Included observations: 365

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.134550	0.021279	-53.31895	0.0000
@TREND("1990M10")	0.002528	0.000101	24.98511	0.0000

R-squared	0.632314	Mean dependent var	-0.674446
Adjusted R-squared	0.631301	S.D. dependent var	0.335440
S.E. of regression	0.203681	Akaike info criterion	-0.339059
Sum squared resid	15.05940	Schwarz criterion	-0.317690
Log likelihood	63.87832	Hannan-Quinn criter.	-0.330567
F-statistic	624.2558	Durbin-Watson stat	0.052996
Prob(F-statistic)	0.000000		

Kwiatkowski-Phillips-Schmidt-Shin test at 1st difference for LSTU

Null Hypothesis: D(LSTU) is stationary
Exogenous: Constant, Linear Trend
Bandwidth: 6 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.049351
Asymptotic critical values*:	1% level	0.216000
	5% level	0.146000
	10% level	0.119000

(Kwiatkowski et al., 1992) (Table 1)

Residual variance (no correction)	0.002192
HAC corrected variance (Bartlett kernel)	0.002825

KPSS Test Equation
Dependent Variable: D(LSTU)
Method: Least Squares
Date: 04/04/21 Time: 18:00
Sample (adjusted): 1990M11 2021M02
Included observations: 364 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003350	0.004931	0.679292	0.4974
@TREND("1990M10")	6.37E-07	2.34E-05	0.027196	0.9783

R-squared	0.000002	Mean dependent var	0.003466
Adjusted R-squared	-0.002760	S.D. dependent var	0.046880
S.E. of regression	0.046944	Akaike info criterion	-3.274231

Sum squared resid	0.797763	Schwarz criterion	-3.252818
Log likelihood	597.9101	Hannan-Quinn criter.	-3.265720
F-statistic	0.000740	Durbin-Watson stat	1.844201
Prob(F-statistic)	0.978318		

Detailed results of Cointegration test

Unrestricted Cointegration Rank Test (**Trace**)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.079177	47.96070	35.19275	**0.0013**
At most 1	0.029322	18.26529	20.26184	0.0920
At most 2	0.020758	7.551472	9.164546	0.1002
Trace test indicates 1 cointegrating eqn(s) at the 0.05 level				
Unrestricted Cointegration Rank Test (**Maximum Eigenvalue**)				
Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.079177	29.69541	22.29962	**0.0039**
At most 1	0.029322	10.71382	15.89210	0.2737
At most 2	0.020758	7.551472	9.164546	0.1002
Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level				
Unrestricted Cointegrating Coefficients (normalized by b'*S11*b=I):				
LCAI	LCOIL	LSTU	C	
7.427105	-1.429216	3.322489	3.982121	
-0.289658	-0.015284	-4.625478	-3.529749	
-0.138897	-2.866361	-0.827875	7.706354	
Unrestricted Adjustment Coefficients (alpha):				
D(LCAI)	-0.010664	-0.003444	0.002661	
D(LCOIL)	0.001855	0.004070	0.011533	
D(LSTU)	-0.006132	0.006835	-0.001203	
1 Cointegrating Equation(s):		Log likelihood	1592.558	
Normalized cointegrating coefficients (standard error in parentheses)				
LCAI	LCOIL	LSTU	C	
1.000000	-0.192432	0.447347	0.536161	
	(0.07165)	(0.11763)	(0.21798)	
Adjustment coefficients (standard error in parentheses)				
D(LCAI)	-0.079204			
	(0.01818)			
D(LCOIL)	0.013777			
	(0.03348)			

D(LSTU)	-0.045542		
	(0.01832)		
2 Cointegrating Equation(s):	Log likelihood	1597.915	
Normalized cointegrating coefficients (standard error in parentheses)			
LCAI	LCOIL	LSTU	C
1.000000	0.000000	12.62864	9.678961
		(3.91305)	(3.83540)
0.000000	1.000000	63.30165	47.51173
		(20.3397)	(19.9361)
Adjustment coefficients (standard error in parentheses)			
D(LCAI)	-0.078206	0.015294	
	(0.01815)	(0.00349)	
D(LCOIL)	0.012598	-0.002713	
	(0.03347)	(0.00644)	
D(LSTU)	-0.047521	0.008659	
	(0.01813)	(0.00349)	

Sample: 1991M03 - 2021M02 (360 after adjustments); Series: LCAI LCOIL LSTU; Exogenous series: S_2001M03 S_2009M12; Lags interval (in first differences): 1 to 4.
*: rejection of the hypothesis at the 0.05 level; **: MacKinnon-Haug-Michelis (1999) p-values

Variance decomposition of LCAI, LCOIL and LSTU

Variance Decomposition of LCAI:

Period	S.E.	LCAI	LCOIL	LSTU
1	0.046456	100.0000	0.000000	0.000000
2	0.083525	99.94639	0.041193	0.012420
3	0.112184	99.84534	0.147196	0.007461
4	0.134437	99.56180	0.269335	0.168869
5	0.152756	98.73726	0.272515	0.990230
6	0.167698	97.55060	0.268638	2.180758
7	0.179966	96.23897	0.304013	3.457013
8	0.190161	94.80499	0.410412	4.784595
9	0.198598	93.26186	0.601635	6.136509
10	0.205602	91.68557	0.864117	7.450308
11	0.211522	90.13329	1.172611	8.694095
12	0.216619	88.64393	1.510022	9.846049
13	0.221083	87.23430	1.869525	10.89617
14	0.225066	85.90274	2.246328	11.85093
15	0.228684	84.64220	2.634494	12.72331
16	0.232020	83.44789	3.027639	13.52447
17	0.235139	82.31636	3.420222	14.26342
18	0.238088	81.24370	3.808204	14.94809
19	0.240905	80.22549	4.188992	15.58552
20	0.243617	79.25718	4.560914	16.18191
21	0.246246	78.33450	4.922805	16.74269

22	0.248809	77.45368	5.273876	17.27245
23	0.251317	76.61142	5.613689	17.77489
24	0.253779	75.80480	5.942109	18.25309

Variance Decomposition of LCOIL:

Period	S.E.	LCAI	LCOIL	LSTU
1	0.085539	4.089026	95.91097	0.000000
2	0.140606	7.812824	92.04255	0.144621
3	0.179956	9.642506	90.08753	0.269963
4	0.208597	10.59938	88.93919	0.461432
5	0.230782	12.16935	87.14554	0.685110
6	0.250037	13.84669	85.37677	0.776537
7	0.267983	15.06529	84.13123	0.803473
8	0.285059	15.74196	83.47678	0.781252
9	0.301222	16.01342	83.25764	0.728940
10	0.316445	16.00919	83.31741	0.673397
11	0.330833	15.84363	83.53197	0.624396
12	0.344537	15.58847	83.82879	0.582740
13	0.357675	15.27894	84.17349	0.547568
14	0.370327	14.93796	84.54410	0.517938
15	0.382543	14.58470	84.92242	0.492882
16	0.394361	14.23311	85.29524	0.471656
17	0.405817	13.89183	85.65444	0.453732
18	0.416946	13.56581	85.99557	0.438621
19	0.427779	13.25750	86.31666	0.425844
20	0.438338	12.96777	86.61725	0.414975
21	0.448645	12.69671	86.89763	0.405665
22	0.458718	12.44384	87.15854	0.397623
23	0.468571	12.20836	87.40102	0.390617
24	0.478218	11.98923	87.62631	0.384458

Variance Decomposition of LSTU:

Period	S.E.	LCAI	LCOIL	LSTU
1	0.046811	0.889011	0.346705	98.76428
2	0.067701	1.244729	0.186967	98.56830
3	0.083974	2.492416	0.121531	97.38605
4	0.099143	5.228594	0.097589	94.67382
5	0.113053	7.423664	0.103058	92.47328
6	0.125623	9.392259	0.096381	90.51136
7	0.137694	11.29726	0.080965	88.62178
8	0.149364	13.13462	0.071091	86.79428
9	0.160547	14.84277	0.067747	85.08949
10	0.171348	16.40131	0.067460	83.53123
11	0.181818	17.78288	0.068505	82.14862
12	0.191950	18.98643	0.070441	80.94313
13	0.201749	20.03558	0.073107	79.89131
14	0.211232	20.95298	0.076228	78.97080

15	0.220407	21.75610	0.079419	78.16448
16	0.229286	22.45979	0.082411	77.45780
17	0.237888	23.07735	0.085121	76.83753
18	0.246227	23.62066	0.087552	76.29179
19	0.254319	24.10045	0.089729	75.80982
20	0.262179	24.52606	0.091674	75.38226
21	0.269823	24.90535	0.093407	75.00125
22	0.277263	25.24485	0.094948	74.66020
23	0.284514	25.55008	0.096318	74.35360
24	0.291587	25.82570	0.097541	74.07676
Cholesky Ordering: LCAI LCOIL LSTU				

…

Author's Statement:

I hereby declare that, in accordance with article 8 of Law 1599/1986 and article 2.4.6 par. 3 of Law 1256/1982, this thesis/dissertation is solely a product of personal work and does not infringe any intellectual property rights of third parties and is not the product of a partial or total plagiarism, and the sources used are strictly limited to the bibliographic references.