ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Master Thesis Financial Economics

The ENSO-effect on U.S. wheat price dynamics:

A short-run and long-run cointegration approach

Author: A.M. Dolk
Student number: 350504
Thesis supervisor: Dr. J.J.G. Lemmen
Thesis co-reader: Dr. Sjoerd van den Hauwe
Date final version: 02-07-2020

ABSTRACT

Based on the theory of storage equation from Fama and French (1987), in this paper I estimate two ARDL models without constant and without trend transformed into case 1 error-correction models to investigate the short-term and long-term direct and indirect effects of the two El-Niño Southern Oscillation (ENSO) phases on the price dynamics of U.S. wheat. The price-effect model comprises of monthly data from April 2006 until November 2019 and investigates the direct ENSO-effects on U.S. wheat price dynamics. The net long position-effect model data ranges from April 1995 until November 2019 and assesses the indirect ENSO-effects on U.S. wheat price dynamics. Both models are controlled for exogenous shocks like inflation and GDP growth and for time-fixed effects as the 2007-2009 financial crisis in the United States of America. This research contributes to existing financial-economic literature by evidencing interdisciplinary causal relationships of meteorological, climatic and financial disciplines in one estimation model with the purpose to provide useful insights for decision-making processes and policy-making by professionals in the financial-, social- and political-economic fields worldwide. The results of both models prove that the El Niño phase and La Niña phase of the ENSO-cycle manifest direct and indirect significant negative effects on the U.S. wheat price dynamics in the long-run and significant weakly positive effects in the short-run.

Keywords: ENSO, climate, El Niño, La Niña, wheat, futures, storage costs, ARDL, ECM.

JEL Classification: B26, B27, G32, Q17, Q54

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

1 Introduction

The weather is hard to predict. Though with today's advanced technology reasonably accurate predictions can be realised. Most of these predictions are immediately absorbed by the agricultural commodity market and hence priced in as the weather accounts for an important factor in crop yields. However, deviations from these predictions cause price fluctuations as the market has to adapt to the new information. This new information is not only important for the farmers and their expected income. Local governments may also be concerned to know if the crop yield will be enough to ensure food security for their people. In addition, on a more financial and international level this information is valuable to the commodity trading industry for determining long and short positions and managing global demand and supply.

In the last few years meteorological models have become more sophisticated, consequently many scientists nowadays argue they can reliably predict major weather and climate events up until six months in the future (Ludescher et al., 2013). The possible prediction of unusual weather events can have a major effect on essential climate, social and economic parameters. The El Niño-Southern Oscillation (ENSO) cycle in particular is one of the largest recurring climate events existing since before humanity and will continue to exist in the future. Due to its immense shape and remotely manifested effects this climate anomaly is repeatedly linked to a wide array of other events not only weather and climate related like tropical cyclones, forest fires and floods, but also related to social and economic problems like distinct disease patterns, social conflicts, inflation, and fishery and crop failure. The latter problem in particular is crucial to all three aforementioned parameters. Crops like wheat, corn and soybeans form the raw materials for one out of two first order needs in the world: food. If crops fail due to climate events, ceteris paribus, prices rise due to scarcity and the perishable nature of crops. The poorest countries will be the first to experience the consequences as their governments no longer have enough funds to import the minimally needed amount to produce sufficient food for its inhabitants. When this happens and food stocks are exhausted, problems like hunger, diseases and social unrest will manifest, pressurising the local economies and politics. Since the ENSO-cycle weather consequences spread worldwide these effects can manifest in several countries or whole regions simultaneously or successively as it takes time for weather conditions to spread. This can affect the global economy and political relations severely.

Current world trade in agriculture accounts for over US$1 trillion annually. Hence, empirical proof and understanding of a significant causal relation between ENSO and wheat prices is invaluable to numerous people, governments and countries worldwide. Wheat production and stocks can be smoothed to prevent extreme price fluctuations and scarcity, providing more food security and political stability to local economies and humanity.

Given the extensive impacts of ENSO on a geographical, social- and financial-economic scale and the importance of wheat crop security and prices for humanity in general, the purpose of this research is to provide useful insights regarding the relation between ENSO-cycle weather events and wheat prices for professionals in many disciplinary fields of expertise to take into account for their decision- and policy-making processes. Although it is not common practice to combine several distinctly different fields of expertise like meteorology, climatology and financial economics within a single research paper, this paper combines the knowledge of these different sciences into one empirical study. Through introducing climatic variables concerning the ENSO-cycle intensity and ENSO-phases to the more financially oriented theory of storage model, controlled for exogenous shocks, I aim to answer the following research question:

Do ENSO-cycle weather shocks have a significant impact on the price dynamics of U.S. wheat?

The results of this research suggest a direct and indirect negative effect of El Niño and La Niña on U.S. wheat price dynamics in the long-run and a significant but weaker positive effect of both phases in the short-run. As the commodity base is defined as the difference between the futures and spot prices of wheat, this result is according to expectations and makes sense because the negative long-run effect implies that information absorbed by the market increases the convergence between futures- and spot prices. In contrast, the weak positive effects of both ENSO-phases on the commodity base implies a short-term divergence of prices due to new information becoming available to the market. The same results are obtained for the indirect effects of the ENSO-cycle on the price dynamics through the non-commercial net long position as a proxy for market sentiment, which in turn is commonly used in literature as a reliable predictor for the price fluctuations of the underlying commodities. Hence, the long-term negative effect suggests a tendency towards a neutral position or net short position held by the non-commercial market participants on the global wheat market, depending on the values of the other factors. The former is in line with stable prices where all available market information is priced in, while the latter would imply a market sentiment in expectation of declining prices in the future. In contrast, the weak positive effect in the short-run implies a tendency towards a net long position. Given the fact this only concerns the short-term effect, this could either mean a market sentiment in anticipation of shortage of supply or increasing prices in the future for other reasons, or a moment of insecurity where market participants want to offset held short-contracts to minimise losses.

The remainder of this paper is organised as follows. The next chapter entails a thorough explanation of the ENSO-cycle climate events, an elaboration on U.S. wheat and the theoretical concepts used in this research. Chapter 3 reviews existing literature regarding similar research. Chapter 4 explains the methodology used for this empirical study. Chapter 5 outlines all details regarding the data used for this research. In chapter 6 I discuss the results obtained from the regression analyses. Followed by chapter 7 providing a validation of these results through a discussion of the robustness checks. At last, in chapter 8 I conclude on my findings by answering this paper's problem statement and discuss the limitations and suggestions for further research.

2 Theoretical background

International trade is essential for economic development, a vital source of foreign exchange earnings and a critical component of food security (FAO, 2017). Particularly in the case of wheat, a globally cultivated and traded agricultural product forming the raw material for our first order need: food. As such, all factors that can influence the quality, quantity, price or productive yield of wheat are of great concern to governments worldwide and humanity in general in order to safeguard the above-mentioned pillars of society. An accurate understanding of one of the world's greatest known anomalous climatic phenomena, the El Niño-Southern Oscillation (ENSO) cycle, is thus essential to the proper assessment of its impact on economically relevant outcomes, such as agricultural output, economic growth or health.

2.1 The ENSO-cycle

The term ENSO-cycle describes the whole inter-annual ocean-atmosphere coupled phenomenon that encompasses the warm El Niño phase, the cold La Niña phase, the atmospheric Southern Oscillation interaction component and the 'normal' phase that exists in between the warm and cold phases (Cane, 1983; Webster and Palmer, 1997; Trenberth and Stepaniak, 2001). However, despite the numerous papers devoted to the subject, there is no universally accepted definition yet (Kug and Ham, 2011). The phenomenon is often incorrectly referred to and understood as El Niño, while technically this term only describes one of three phases of the whole cycle. Both the meaning of the term El Niño as the climatic event itself have evolved over time (Trenberth, 1997; Trenberth and Stepaniak, 2001). To begin with, a distinction should be made between the terms 'El Niño', its counterpart 'La Niña', the Southern Oscillation component and the El Niño-Southern Oscillation (ENSO)-cycle as a whole, based on their distinguishing characteristics (see section 2.1.1-4). Then there is a discrepancy in the scientifically used (formal) term El Niño that does make that distinction and the publically used (informal) term El Niño that does not draw any distinction at all (Trenberth, 1997). Furthermore, a separation of beliefs has emerged in recent years leading to an ongoing discussion among scientists whether there is only one El Niño event, commonly known as the 'conventional' El Niño[1], or two distinctly different El Niño events. The other El Niño event would be the newly discovered El Niño Modoki[2,3], which supposedly has different characteristics and hence different impacts than the conventional El Niño (Yeh et al., 2009). To prevent confusion, this research only uses terms and referrals of scientific nature and focuses on the effects of the La Niña phase, the 'conventional' El Niño phase and the ENSO-cycle as a whole.

The ENSO-cycle appears in the equatorial Pacific Ocean and its surroundings. For perspective comprehension, according to the National Oceanic and Atmospheric Administration (NOAA) the Pacific Ocean is the largest and deepest ocean basin on Earth, covering over 155 square kilometres, equals roughly thirty percent of the Earth's surface and holds more than half of the Earth's open water supply (NOAA,

---

[1] Also called or referred to as 'Canonical' El Niño (Cane, 1983; Takahashi et al., 2011), Eastern Pacific (EP) El Niño (Gu et al., 2015; Kao and Yu, 2009; Karnauskas, 2013; Ren et al., 2016; Hu et al., 2016), or Cold Tongue (CT) El Niño (Kug et al., 2009).
[2] Also called or referred to as 'Dateline' El Niño (Yeh et al., 2009; Karnauskas, 2013), Central Pacific (CP) El Niño (Kao and Yu, 2009; Hu et al., 2016; Ren et al., 2016), or Warm Pool El Niño (Kug et al., 2009).
[3] "Modoki" is a classical Japanese word meaning "a similar but different thing" (Ashok et al., 2007).

2015). Moreover, the tropical Pacific Ocean is characterised by warm surface water in the west (i.e. around 29–30 degrees Celsius) and much cooler water in the east (i.e. around 22–24 °C) according to Webster and Palmer (1997). Hence, when normal weather and water conditions in the Pacific Ocean region are distorted, the impacts will manifest not only on a local but also on a global scale (Webster and Palmer, 1997).

2.1.1 El Niño

Starting with the very basics, '*el niño*' is Spanish for 'the little boy' or 'child'. When written with capital initials the term '*El Niño*' originally refers to the Christ child Jesus (Trenberth, 1997; Wang and Fiedler, 2006). The latter is the name Peruvian sailors and fishermen originally gave to an annual, weak warm ocean current (i.e. la corriente del Niño) in the Pacific Ocean which runs southward along the coast of Peru and Ecuador (Holton and Dmowska, 1989; Pidwirny, 2006).[4] The reason for this name is due to its occurrence around Christmastime (Philander, 1989). However, in the late 19th century Peruvian geographers first discovered that every few years this annual weak warm ocean current is significantly warmer than other years, extending far more south and is accompanied by other unusual oceanic and atmospheric phenomena (Philander et al., 1989; Wang and Fiedler, 2006). Due to this discovery and other foreign-based scientific expeditions in the early 20th century[5], the concept of El Niño became known within the world's scientific community mainly as referring to these occasional abnormal conditions rather than the annual occurrence, which was forgotten (Wang and Fiedler, 2006). Subsequently, other oceanographers realised around the 1960s that this occasional extreme coastal warming also extends immensely far offshore into the Pacific Ocean, changing both local and regional ecology (Philander et al., 1989; Trenberth, 1997). Based on this knowledge a discussion started about which oceanic and or atmospheric condition(s) cause(s) this particular phenomenon; what are the directly resulting conditions, and if any structural pattern could be identified. A discussion remaining relevant in today's literature and that I will elaborate on in section 2.1.2.

Referring to an essential characteristic of the Pacific Ocean, under normal conditions, the warm surface water in the west is usually referred to as the Pacific "warm pool" (Webster and Palmer, 1997). This warm pool annually exports part of its heat. The inability of the warm pool to export enough heat each year naturally builds up its heat content, resulting in an outburst of this heat every few years possibly triggered by supporting atmospheric conditions (Wyrtki, 1975). Cane (1983) found the above series of events, which are generally assumed to respond to meteorological parameters, to be characteristic for the 'Canonical' El Niño and rhythmic in decadal timeframes, hence the anomaly's close link to the annual cycle and oceanic conditions. Publicly known as the conventional type of El Niño (or warm phase). These theories from Webster and Palmer, Wyrtki and Cane also seem to explain the earlier discovered annual weak warm ocean current along the coast of Peru. Although being part of the earliest theories[6], the assumption of ocean-atmosphere coupling to explain El Niño remains generally accepted nowadays (Karnauskas, 2013). However, the extraordinary 1982-83 and 1997-98 El Niño events showed different evolutions of conditions with respect to this "canonical" composite according to Takahashi et al. (2011). These two notable events

[4] For a formal definition as it should appear in a dictionary according to Trenberth (1997), see the work of Glantz (1996).
[5] See also the works of Murphy (1926) and Lobell (1942).
[6] See also the work of Bjerknes (1969).

in particular and the further evolution of El Niño-tied climatic conditions in more recent years has led to the next discussion currently at hand, concerning whether or not there exists a second and distinctly different El Niño, called the El Niño Modoki. Amongst others, Karnauskas (2013) and Takahashi et al. (2011) argue that the conventional 'canonical' El Niño shows a non-linear evolution over time, one which scientists perhaps cannot fully grasp within a single empirical model yet. In contrast, Ashok et al. (2007), Kao and Yu (2009) and Ren and Jin (2011) suggest the existence of different versions of El Niño and believe El Niño Modoki to be a distinguishably separate event. As several different techniques and indices can be used to measure atmosphere and ocean conditions, it is a delicate issue to choose which index to use to indicate the possible occurrence of an El Niño. Consequently, it is difficult to take a firm stance for either strand of literature. For the sake of applicability and public understanding, in this research I assume the existence of only the conventional type of El Niño.

2.1.2 The Southern Oscillation

The term Southern Oscillation describes a rhythmic variation in sea-level air pressure (i.e. oscillatory exchange of mass) influencing the (reversal of) atmospheric circulation over the tropical Pacific Ocean (Pidwirny, 2006; Julian and Chervin, 1978). This natural mode of oscillation has two complementary phases, one which causes ocean-atmosphere conditions known as El Niño (the warm phase) and the other causes the conditions known as La Niña (the cold phase) (Philander, 1989; Fedorov and Philander, 2000). These two phases are interspersed with a rather instable but neutral state (Webster and Palmer, 1997).

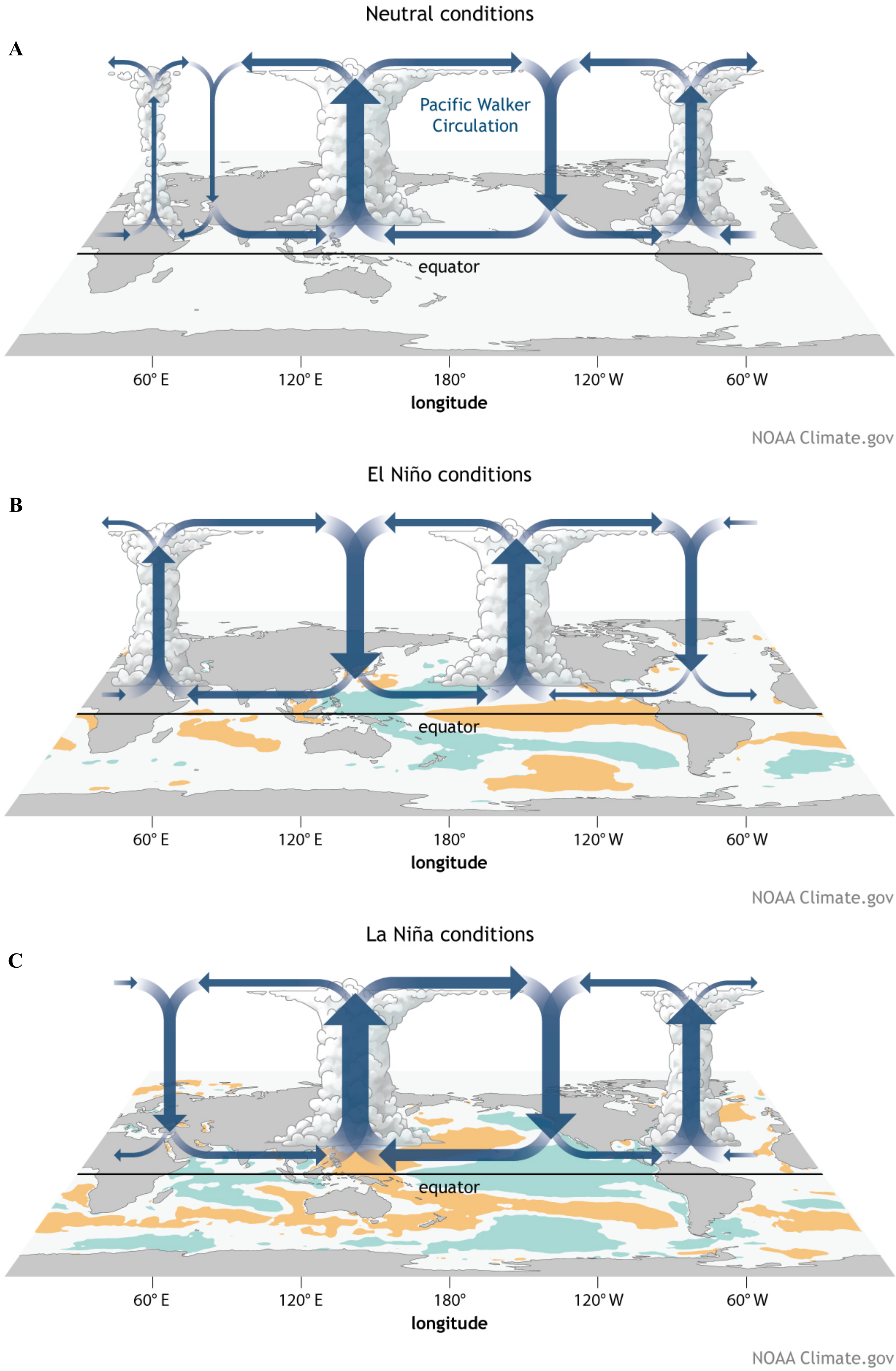
Sir Gilbert Walker was the first to demonstrate the existence of irregular inter-annual fluctuations in the tropical and global atmospheric circulation, involving major changes in rainfall patterns and wind fields over the tropical Pacific and Indian oceans, which he named the Southern Oscillation (Philander, 1989; Walker and Bliss, 1932). Ever since, extensive research has been conducted with divergent objectives, from trying to understand weather patterns to forecasting possible agricultural and economic consequences of major climate events like the El Niño-Southern Oscillation.

Cashin et al. (2015) argue that in 'normal' years a high-pressure system develops at surface level over the coast of Peru and a low-pressure system builds up in northern Australia and Indonesia, called the Pacific Walker Circulation[7] (see Fig. 1-A). As a result, trade winds[8] move strongly from east to west over the Pacific Ocean surface, carrying the warm surface waters westward. In turn, this movement brings precipitation to Indonesia and Australia. Simultaneously, cold nutrient-rich water wells up to the surface along the coast of Peru, boosting the fishing industry in South America. During El Niño and La Niña years, this system is 'distorted', resulting in both atmospheric and water conditions to behave differently. This generally accepted conception is also affirmed by other scientists and the NOAA.

[7] The Pacific Walker Circulation is a large atmospheric overturning cell spanning the tropical Pacific Ocean, characterised by rising air motion (lower sea-level pressure) over Indonesia and sinking motion (higher sea level-pressure) over the eastern Pacific (L'Heureux et al., 2013).

[8] Trade winds are surface winds generally dominating the airflow in the tropics. These winds blow from about 30° North and South latitude to the equator. Trade winds in the Northern Hemisphere have northeast to southwest direction and are often referred to as the Northeast Trades. Southern Hemisphere trade winds have southeast to northwest direction and are called Southeast Trades (Pidwirny, 2006).

Figure 1: The Pacific Walker Circulation under the different ENSO-cycle conditions.



In El Niño years, the Southern Oscillation causes the atmospheric circulation above the tropical Pacific Ocean to reverse (see Fig.1-B – where the orange areas represent ocean warming and blue-green areas represent ocean cooling). As the 'normal' low pressure system in the western Pacific turns into a high-pressure system because of the accumulated pool of warm ocean water, Cashin et al. (2015) and many others[9] claim that the air pressure drops along the coast of South America and over large areas of the central Pacific, also partly due to the upwelling of the cold deep ocean water in the preceding neutral period. They continue that this shift causes the trade winds to weaken and allowing the equatorial counter current (which flows from west to east) to accumulate warm ocean water along the coastlines of Peru. In response to the latter, the thermocline[10] drops in the eastern part of the Pacific Ocean, cutting off the upwelling of cold water along the coast of Peru. In terms of weather conditions, these reversed conditions bring drought to western Pacific and Australia, heavy rains and floods to the equatorial coast of South America and hurricanes and stormy weather to the central Pacific.

In La Niña years, mostly directly following an El Niño period, the 'normal' atmospheric conditions go into overdrive to compensate for the reversed conditions in the preceding period (see Fig.1-C). The trade winds intensify dramatically, shifting all the warm ocean water back to the west resulting in even higher sea surface temperatures than usual (Holton, Dmowska et al., 1989). Subsequently, the low-pressure system forms above northern Australia and Indonesia where it causes excessive rain fall, while in the central and eastern Pacific an abnormal accumulation of cold water occurs as the thermocline resurfaces (Fedorov and Philander, 2000; Pidwirny, 2006). The latter causes a surface high-pressure system to develop again along the coast of Peru that brings drier but milder winters to the South American region and colder winters to North American regions (Rasmusson and Wallace, 1983).

2.1.3 La Niña

Similar to its counterpart, the term '*la niña*' is Spanish for 'the little girl'. When written with capital initials the term '*La Niña*' refers to the opposite climatic condition of El Niño, also known as the 'cold phase' (Trenberth, 1997). This cold phase thanks its name to the colder than average sea surface temperatures in the central and eastern Pacific, which are not only responsible for extremely cold winters in Northern America and Canada, but on the other hand also for the return of a booming fishing industry along the coast of Peru and Chile (Badjeck et al., 2010). La Niña phases are also associated with disturbed global climate conditions, which in some cases can prove to be far more disastrous than the better-known El Niño conditions. Ranging from extreme rainfall and floods in the far western Pacific area, tropical storms and hurricanes in various parts of the world, to severe drought in eastern and southern Africa as well as in South American countries (Holton, Dmowska et al., 1989). Despite its serious impacts on global climate and weather patterns, La Niña is often wrongly seen as El Niño's "little sister". Unfortunately, this results in little explicit literature about this climate phenomenon on its own, especially in relation to its impacts on

[9] Julian and Chervin (1978), Cane (1983), Philander (1989) and Webster and Palmer (1997) agree with this line of thought.
[10] The thermocline is a body of ocean water where the greatest vertical change in temperature occurs. This boundary is usually the transition zone between the mixed layer of warm ocean water near the surface (that is often much influenced by atmospheric fluxes according to Cashin et al., (2015)) and the cold deep-ocean water layer (Pidwirny, 2006).

global social and economic parameters. However, the characteristics and consequences currently known of La Niña are broadly accepted within the scientific literature.

2.1.4 The ENSO-cycle and its effects on global agriculture

Following the seasonal cycle, the ENSO-cycle accounts for the second most important source of global weather and short-term climate variation (Glantz, 2001; Rosenzweig et al., 2001; Goddard and Dilley, 2005). The El Niño-Southern Oscillation is mainly a cyclical event due to the Southern Oscillation component that is a rhythmic climate fluctuation, causing the extreme conditions known as El Niño and La Niña to reoccur roughly every 3 to 7 years respectively according to Webster and Palmer (1997). The timeframe of this pattern proves to vary over time and some even warn that both the frequency as the degree of extremity of ENSO events will increase dramatically due to greenhouse warming (Cai et al., 2014). Moreover, the NOAA claims that each phase lasts for approximately 9 to 12 months on average with exceptions of strong prolonged El Niño and La Niña episodes that may last for up to 2 years. In between these two extremes the duration of the quasi-equilibrium state depends on (a series of) random disturbances of the background climate state that both contribute to the maintenance and disturbances of this instable neutral phase (Webster and Palmer, 1997; Fedorov and Philander, 2000). The complementary phases of the ENSO-cycle directly impact economies of fishery dependent and crop producing countries surrounding the Pacific Ocean. Indirectly, the ENSO-cycle also impacts other countries' economies all around the globe through its atmospheric teleconnections[11] that affect precipitation patterns worldwide (Abdolrahimi, 2016).

The biggest caveat in defining cause-effect relationships over time is the fact that the explanation for El Niño and La Niña involves a circular argument as it constitutes of a two-way interaction between the ocean (i.e. El Niño and La Niña) and the atmosphere (i.e. the Southern Oscillation) according to Cane (1983) and Abdolrahimi (2016).

Consequently, Fedorov and Philander (2000) argue that changes in sea surface temperatures are both the cause and consequence of wind fluctuations, inherently these interactions can amount to positive and negative feedback processes. A growing general acceptance implies a combination of the measurement index for the atmospheric air-pressure differentials of the Southern Oscillation and the measurement index for the oceanic temperature gradients should form the key to solving the puzzle as they are the most widely used indices to determine the presence and magnitude of an El Niño and La Niña phase (Cashin et al., 2015; Abdolrahimi, 2016). In particular, because the Southern Oscillation Index (SOI) and the Sea Surface Temperature (SST) gradients are exceptionally strongly correlated with the connected weather variations and resulting disruptions in precipitation, crop production and fishery to name a few.

Now a comprehensive understanding of the ENSO-cycle related phases and their manifestations is established, it follows logically that an agricultural product like wheat is particularly vulnerable for these kinds of changes in weather and climate conditions. As the cereal usually spans a sowing and harvest period of seven to eight months for winter wheat and around four months for spring wheat and durum wheat, it is

[11] A teleconnection is a spatially and temporally large-scale anomaly that can influence the variability of the atmospheric circulation on both a local and global scale (NOAA, 2017).

subject to a long period of weather dependent growth. Distortions like extreme drought or excessive precipitation can have disastrous consequences on the quality and quantity of wheat harvested at the end of the cultivation period. In turn, this can have major influences on food security, foreign exchange income from export and inflation. According to previous research of Abdolrahimi (2016) and others, the ENSO-cycle in particular mainly affects agricultural output in either of below listed three ways:

I. Effect on crop production (quality) due to disturbing influences of ENSO on temperature and precipitation variations (Naylor et al., 2002);
II. Pest damage due to weather circumstances caused by ENSO that provide conditions for the growth of fungi and insects (Rosenzweig et al., 2000);
III. Total crop failure due to hazardous weather conditions like severe drought, flooding and storms caused by ENSO conditions (Changnon, 1999).

2.1.5 Quantification of the ENSO-cycle conditions

In relation to the ENSO-cycle, there are several ways through which the air pressure, temperature, winds and water conditions are indexed to track in which phase of the natural oscillation it is currently in. The most well-known indices are the following:

- **Southern Oscillation Index (SOI):** a standardised index based on barometric sea level air pressure differences observed between two observation stations in Tahiti and Darwin, Australia;
- **Oceanic Niño Index (ONI):** this index defines a(n) El Niño (La Niña) as characterised by a minimum of five consecutive 3-month running mean of Sea Surface Temperature (SST) anomalies in the Niño-3.4 region that is above (below) the threshold of +0.5°C (-0.5°C);
- **Trans-Niño Index (TNI):** this index is given by the difference in normalised SST anomalies between the Niño-1+2 regions combined and the Niño-4 region.

The latter is a relatively new index suggested by Trenberth and Stepaniak in 2001 in an effort to regain the ability to describe the earlier discussed non-linear evolution of El Niño and its diversity of patterns after the extraordinary and deviating El Niño events in 1982-83 and 1997-98 (Takahashi et al., 2011). However, the standardised SOI and ONI indices are by far the most commonly used indices for both environmental and socio-economic research, also after 2001.

2.2 Wheat

Wheat is a cereal grain commonly believed to be first cultivated in the Levant region and the Ethiopian Highlands, dating back to at least 8000 years ago. It has always been the most important source of food grain for humans and this continues nowadays (Curtis, Rajaram and Macpherson, 2002). Currently, wheat is grown on more land area than any other commercial crop and reached a record production in 2016 of over

750 million metric tonnes, according to 2017 data of the United States Department of Agriculture (USDA). Together with maize and rice, wheat is one of the most widely produced cereal crops on Earth, cultivated on more than 200 million hectares of land in over 124 countries (Abdolrahimi, 2016). The wheat cereal is also estimated to account for more than 20 percent of calories and on average 30 percent of proteins consumed by humans according to statistics from the Food and Agriculture Organisation (FAOSTAT) and the National Association of Wheat Growers (NAWG) (2017) with respect to the USDA's dietary reference intake (DRI) scale. Again, these numbers illustrate the importance of acquiring knowledge of and spreading awareness for the potential impacts ENSO-events can manifest on our global agricultural production and food security.

2.2.1 Wheat contracts

As wheat is a physical and agricultural product, its market price mechanism is far more complex compared to non-physical products like financial assets (e.g. stocks and bonds). A contract for the sale of a financial asset, where ownership of the asset transfers from seller to buyer, is often settled within nano-seconds and hardly has any transaction costs. Prices of financial assets are mainly influenced by the well-known concepts of supply-and-demand, (inter-)national monetary policies, interest rate, profitability and credit rating of the partially-owned underlying asset (e.g. a company), and uncertainty.

In contrast, settlement (i.e. delivery) of a contract for the sale of wheat can take up to a few days at best or longer and knows significant transaction costs. Therefore, the price-setting mechanism of wheat is totally different, mainly because of storage costs for the actual good, the opportunity cost of capital in storing grains and the transport costs for delivery of the goods, all depending on the contract (delivery) terms agreed on between buyer and seller. Needless to say, storage and transport of physical goods embody an entire market of supply-and-demand in themselves. Prices of wheat are hence influenced by a wide variety of factors in addition to those influencing financial assets and the uncertainties associated with the delivery of wheat are far greater. Technically speaking, there are three types of wheat contracts as I explain below:

Spot contract

A spot contract is a contract that implicates the purchase or sale of a commodity for immediate delivery and payment on the spot date, which is usually up to two business days after the trade date. However, in the international commodity trading business of dry bulk like wheat, delivery within two business days is close to impossible and therefore highly unusual. Hence, a spot contract is understood as a forward or futures contract for delivery in the current month. Since the latter is no common practice, such a spot contract may not exist or may not be actively traded for every commodity and or in every month (Pindyck, 2001). Nonetheless, it is essential to rewrite the above in a basic formula:

$$S = X \tag{1}$$

The spot price S, or spot rate, is the current price X (in U.S. Dollars per bushel[12]) of the asset quoted for immediate settlement of the spot contract. Meaning direct delivery of the actual goods.

Forward contract

A forward contract is a bilateral contract that involves an agreement of contract terms (i.e. price, quantity and commodity specifications) on the current date t for the purchase or sale of the underlying physical asset, delivery of this commodity and payment at a specified maturity date T, for which holds that $(t \leq T)$, or during a specified delivery period (Huisman, 2014). Hence, a forward contract is a form of a derivative since its value (i.e. the forward price, or forward rate) is (partly) dependent on the value (i.e. spot price) of the underlying commodity. Two simplified versions exist of how to derive the forward price of a financial asset that provides no income and has no storage costs, using discrete compounding[13] and continuous compounding[14] respectively:

$$F_0 = S_0(1+r)^T \tag{2}$$

$$F_0 = S_0 e^{rT} \tag{3}$$

Here, F_0 denotes the forward price at $t = 0$ for the forward contract deliverable in T amount of time (e.g. years or months). Next, S_0 denotes the spot price of the same asset at $t = 0$ and r denotes the risk-free rate assessed over the period of time T until maturity or delivery of the contract. Note that I assume no arbitrage opportunities[15] here, meaning the Martingale Restriction as introduced and explained in further detail by Longstaff (1995) must be satisfied, otherwise these equations would not hold. These simplified formulas are particularly fit for financial assets, as the price of physical assets would consist of multiple factors.

Futures contract

A futures contract is a standardised (forward) contract that is traded on an exchange (e.g. the Chicago Mercantile Exchange). As opposed to a normal forward contract, that is not standardised and traded directly between a buyer and seller. The standardisation of the contract implies that price, quantity, commodity specification, maturity or expiration date and delivery date or procedure are all predetermined in the contract. Next, exchanges and the associated clearing system impose a margining system that applies to every trader. Before trading, a trader has to open a margin account at a clearing bank and has to put an

[12] A ‘Bushel’ is the most commonly used unit of weight for dry bulk commodities in the U.S. and surrounding countries, which equals roughly 60 pounds. See Appendix A for the weight conversion table.

[13] With discrete compounding the forgone interest on the principal is in theory assessed and added to the value of the principal, in this case the spot price, of the asset at a finite set of times, e.g. (semi)-annually, until maturity to correct for the lost opportunity to wait. See also Broadie et al. (1999) for a more thorough mathematical explanation of this concept.

[14] With continuous compounding the forgone interest on the principal is in theory assessed and added to the value of the principal, in this case the spot price, of the asset in infinitely small increments of time until maturity to correct for the lost opportunity to wait. In practice this is impossible so more common increments are used like days, weeks and months. See also Broadie et al. (1999) for a more thorough mathematical explanation of this concept.

[15] An arbitrage opportunity arises shortly when one can make a guaranteed profit without investing capital or bearing risk due to a price differential between two markets (Dybvig and Ross, 1989).

amount of cash on the account (usually a percentage of the contract value). The money on the account is then used to settle the daily changes in the market value of the contract (marking to market). When the money on the account reaches a lower bound, the trader receives a margin call and has to put additional funds on the account (Huisman, 2014). Unlike common with financial assets, actual delivery of physical assets often takes place in dry bulk commodity trading.

Important to note is that according to Hull (2013), if the maturity and commodity price are the same, forward and futures prices are assumed to be equal. In case interest rates are uncertain, Hull (2013) claims that forward and futures prices can theoretically differ slightly: a strong positive correlation between the interest rate and the commodity price implies a futures price that is slightly higher than the forward price, and vice versa in case of a strong negative correlation. However, as interest rates are nowadays well-documented with daily, monthly and yearly frequency and cater for plentiful different maturities, I assume no interest rate uncertainty in this research. Consequently, the formulation of the futures price is equal to that of the forward price, meaning formulas (2) and (3) still hold.

The advantages of a futures contract over a forward contract are that it partly reduces counterparty risk and has a higher liquidity in terms of trading due to its standardisation. The disadvantage is that futures contracts are often traded in a predetermined fixed batch of units, unlike bilateral contracts as spots and forwards. For example, the Chicago Mercantile Exchange (CME) rules read that the minimum wheat futures contract[16] allowed by the exchange is equal to 5,000 bushels of wheat for United States (U.S.) contracts and 50 Metric tonnes[17] of wheat for European (EU) contracts. This translates to approximately 136,100 Kilograms and precisely 50,000 Kilograms respectively.

2.2.2 Open interest and Net long positions

Every Friday of the week the U.S. Commodity Futures Trading Commission (CFTC) publishes a report called the CoT, short for Commitment of Traders. This report contains information about the preceding Tuesday of the same week and can be seen as the CFTC's order book of the commodity Futures market in the U.S. As the CoT contains market information of all active players about the previous Tuesday this report is used as a lagging indicator of market sentiment in the futures market. The report divides the market in three categories of traders:

- **I.** Commercial traders (i.e. producers);
- **II.** Non-commercial traders (i.e. large traders like hedge funds, banks and institutions); and
- **III.** Non-reportable traders (i.e. small traders).

The commercial traders are in the market to hedge their products and, as such, are no accurate indicators of market sentiment. Likewise, the non-reportable traders are small traders whose positions are not large

[16] This concerns the minimum of a normal wheat futures contract. However, recently the Chicago Board of Trade (CBOT) introduced 'mini-sized' wheat contracts of 1,000 bushels per contract in an effort to overcome this issue, see chapter 14B of the CME CBOT Rulebook (2017) on Wheat Futures for more detailed information.
[17] A Metric tonne is a common European unit of weight used for dry bulk commodities. See Appendix A for the weight conversion.

enough to indicate or influence any market sentiment. The non-commercial traders are the most important group to derive market sentiment from. Within this group there are two types: high volume traders and speculators. Speculators are in the market for profit, not interested in taking delivery of the underlying commodity. High volume traders are in the market for the actual purchase of the underlying commodities and taking delivery in the future. The CoT contains information on the net positions per category and type of trader and the total open interest of all active players on the futures market.

A net position is the surplus of long and short trades, also explained as the level of overall long versus short contracts. The net position of the non-commercial traders is generally understood as a strong indicator of market sentiment and price trends within the commodity trading industry, as also proven by Hong and Yogo (2012). The open interest is the total number of open contracts, both long and short, between all market participants also including the commercial and non-reportable traders. This is also an indicator of market sentiment but less strong as it also contains hedging positions from wheat producers which are usually opposite to market sentiment.

2.3 Theory of storage versus the expectations theory

As previously mentioned, the foregone formulas do not account for important factors that become relevant when trading physical commodities. In contrast to financial assets, physical assets require actual transport for delivery and physical storage until the maturity date of the contracts or until the producer/seller decides to put the assets up for sale. Therefore, Working (1949) and others[18] introduced the 'theory of storage' which explains the difference between contemporary spot and futures prices in 3 factors: (i) interest forgone in storing a commodity; (ii) warehousing costs, and (iii) a convenience yield on inventory. A different approach by Cootner (1960) and others[19], called the 'expectations theory', splits the futures price into an expected risk premium and a forecasted future spot price. Fama and French (1987) explored these two most common views on physical commodity futures prices and found that the first approach entails a better explanatory forecasting power when the formula is used as a regression, as opposed to the second view that bears some major constraining flaws.[20] Therefore, in this research I will further build on Fama and French's rewritten version of Holbrook Working's original theory of storage equation where the difference between futures and spot prices is assumed to be equal to the interest forgone in storing, the warehousing costs and the convenience yield. I will elaborate in further detail on this equation in the methodology section.

2.3.1 Storage costs and the convenience yield on inventory

The concept of physical storage entails a market of itself with supply, demand, duration and its implied price fluctuations. Storage of the actual physical goods does not only encompass costs for storage, depreciation of the goods and opportunity costs of capital. Inventory of goods (e.g. wheat) also implies either the opportunity to meet unexpected demand spikes on the spot market and reap (excess) returns from the

18 See also the works of Kaldor (1939), Brennan (1958) and Telser (1958).
19 See also the works Dusak (1973), Breeden (1980) and Hazuka (1984).
20 The regression corresponding to the theory of Cootner (1960) and others allocates all abnormalities to the expected premium, including irrational forecasts and measurement errors. Consequently, this approach to the theory of storage is not reliable.

corresponding price hikes (Lautier, 2008), or productive value to use the raw goods as input for production of other commodities (e.g. flour in the case of wheat) (Fama and French, 1987). There are two sorts of storage:

- **Commercial storage:** storage capacity offered as commercial service by companies that do not own the stored goods themselves and aim to maximise profit, mostly located at off-farm sites.
- **Non-commercial storage:** storage capacity owned by the same business that owns the goods, mostly farmers with their own on-farm storage or off-farm cooperative storage sites where farmers store goods for themselves, this cooperative storage organisation aims to maximise benefits of services offered and usually operates at zero-profit.

2.4 The Efficient Market Hypothesis (EMH)

Eugene Fama developed the Efficient Market Hypothesis (EMH) in 1970. The hypothesis states that in an ideal capital market, prices of assets and securities at any point in time fully reflect all available information. A market is then called 'efficient'. This implies that successive price changes are independent. However, a few conditions need to be met by a capital market in order for the prices to be able to efficiently adjust to the available information. These conditions comprise of the following:

I. No transaction costs;
II. All available information is free for every market participant;
III. Everybody agrees on the implications of present information for the current price and distributions of future prices of each security (Fama, 1970).

This research however, is based on the assumption that EMH does not hold. In practice, many capital markets contain inefficiencies. First of all, to ameliorate the cash market inefficiencies, the CME first introduced a Variable Storage Rates (VSR) system in 2010 for the storage market on a number of dry bulk commodities amongst which wheat. In particular, this system serves the purpose to help achieve a better price convergence between the spot and futures prices during the delivery periods of the underlying assets. In addition, transaction costs are present in the form of opportunity costs of capital. On top of that it is common knowledge that humanity is not yet capable of flawlessly predicting future weather events. Hence, there is information that the market is consistently unable to process in the pricing mechanism of the wheat futures market. In trying to explain a (significant) part of the so far un-identified 'uncertainty' within the equation of the theory of storage by ENSO events I implicitly assume that the capital market is not efficient and that prices do not immediately adjust to all publicly available information.

3 Literature review

In this research, I investigate the possible existence of short-term and long-term (causal) relationships between the future-spot price parity of wheat and the respective intensity of the El Niño phase and La Niña phase of the ENSO-cycle. Also, I investigate the same possible correlation for the non-commercial net long position with respect to the warm and cold phases of the ENSO-cycle.

The models I use are based on an extended theoretical framework of the theory of storage provided by Fama and French in 1987 with additional assumptions. Also, I add ENSO-intensity and -phase to the equation as independent variable and dummy variable respectively while controlling for the exogenous influences of inflation, nominal GDP growth and the time fixed effect of the 2007-2009 U.S. financial crisis. In doing so I aim to contribute to existing financial-economic literature through combining meteorological, climatic and financial disciplines to derive meaningful evidence of interdisciplinary causal relationships in one estimation model. The outcome of this research can provide useful insights for decision-making processes and policy-making by professionals in the financial-, social- and political-economic fields worldwide.

Back in 1987 Fama and French find that for 10 out of 21 commodities that they included in their original theory of storage-model, the futures price contained significant forecasting power for the spot price over the sample period January 1967 to May 1984. Stating the futures prices primarily respond to the storage costs, interest forgone in storage and a theoretical marginal convenience yield in storing commodities rather than risk premiums. After Fama and French, Perales (2010) is the only other more recent paper I find to examine the theory of storage, let alone in particular relation to wheat price dynamics. Perales uses a Baba, Engle, Kraft and Kroner-model (BEKK) confirming that the theory of storage from Kaldor (1939) and Fama and French (1987) holds, meaning the storage costs (i.e. interest-adjusted spread between spot and futures prices) has a significant positive effect on the variabilities of spot and futures returns for corn and wheat over the period 1975 to 1999. However, Perales also claims that agricultural commodities' spot and futures prices are stationary I(0), which he motivates would render the Error Correction Model (ECM), previously used by Kroner and Sultan (1993) and Ng and Pirrong (1994) useless. Stating the BEKK-model is therefore a reliable alternative. This might pose a problem in conducting the error-correction form of the ARDL-model, also called the Error Correction Model (ECM). One difference being that Perales uses daily frequency of data whereas I use monthly data just like the vast majority of research uses monthly or quarterly data.

Other than Perales, in the subsequent years following Fama and French's findings a great deal of scientific literature mainly uses other models and focuses on ENSO impacts on more general economic factors like world prices, economic activity, commodity returns, and global production of crops per country and or per crop type. For instance, Podestá et al. (1998) investigate associations between yields of major crops in the Argentine Pampas (central-eastern Argentina) and El Niño–Southern Oscillation (ENSO) phases for the period 1912 to 1990. While for maize, sorghum, soybean and sunflower certain statistically significant causal relationships have been evidenced, they could not find any associations with ENSO-phases for wheat, the only winter crop that was considered in the research. One of the arguments they use to explain this lack of causality is the fact that wheat has the widest geographic distribution of all crops considered in their research, this has implications for wheat's link with the ENSO-cycle phases. They continue that regionally inhomogeneous ENSO effects may cancel out when analysing national-level yields.

A few years later in 2002, Brunner shows that for the period between 1950-2002, the ENSO-cycle has explanatory power accounting for ten to twenty percent of the movements in world consumer price inflation and world economic activity. In particular, he finds positive surprise/shock in ENSO intensity to raise commodity price inflation (in real terms) about 3.5 to 4 percentage points six months after the shock. Then, prices fall by a similar figure in the second and third year after the shock. Especially in the longer run (about sixteen quarters i.e. four years later) ENSO shocks have a much stronger impact on real food and primary commodity prices. In his research, Brunner also claims that the SOI anomaly measure of ENSO intensity has much stronger statistical relationship with the economic variables than the SST anomaly index.

Even Philander states already in (1989) that "interactions between the tropical Pacific Ocean and the atmosphere cause predictable inter-annual fluctuations in climate…" and claims that when eliminating the superimposed high frequency fluctuations by filtering data, it would be possible to make predictions by extrapolating this low-frequency trend. Making use of such a technique Cashin et al. (2015) more recently show the impacts of El Niño shocks on macroeconomic growth, inflation, energy and non-fuel commodity prices on a country-specific level covering thirty-three countries during the 1979Q2-2013Q1 period using quarterly data. They deduct that countries like Indonesia, Chile and South Africa experience a temporal decrease in economic activity in response to an El Niño shock while countries of major economies such as the U.S. and Europe face growth-enhancing effects, evidencing statistically significant effects on real GDP growth. Many countries in their sample can experience a (short-run) price increase up to 5.31% of non-fuel commodities, four quarters after the El Niño shock. Cashin et al. (2015) state this is due to a circular effect, the negatively impacted countries yield less supply of non-fuel commodities while the enhanced GDP growth of positively impacted countries boosts demand for the same commodities. Hence, the initial impact of the El Niño shock echoes in the months and even years to follow. Apart from direct effects on commodity price levels, Cashin et al. also prove that El Niño has an (in)direct influence on yet another economic variable: inflation. This poses an interesting challenge in distinguishing the genuine impact of ENSO-shocks in the research model used in this paper as I include inflation and nominal GDP growth as control variables, which are supposedly also (partly) influenced by ENSO.

Next, Abdolrahimi (2016) finds a statistically significant impact of ENSO on the individual production of wheat from 39 percent of the main wheat-cultivating countries during the 1962-2009 period. She explains that while El Niño shocks cause production to decrease substantially in Bulgaria and Morocco, wheat production in the U.S. and Argentina increases with 4.9 and 11.1 percent respectively. Following these findings of increased supply in the U.S., ceteris paribus, I expect El Niño to have a negative effect on U.S. wheat prices and La Niña to have a positive effect on U.S. wheat prices as it has a strong negative effect on production. Furthermore, Abdolrahimi claims that of the two extreme phases of the ENSO-cycle, the cold La Niña phase appears to be more important in the ENSO-wheat production relationship which confirms my earlier statement that La Niña is often underestimated.

Finally, over the January 1982 to December 2014 period based on monthly data and using a Vector Smooth Transition Autoregressive (VSTAR) modelling framework, Ubilava (2017) shows a significant negative correlation between ENSO-cycles and wheat price cycles for 5 different wheat prices. Amongst which U.S. wheat prices. Ubilava uses the Niño3.4 Index (also known as the ONI-index) as a proxy for the ENSO intensity. He concludes that wheat prices seem to increase after La Niña episodes, and decrease after El Niño episodes. Although a very different model is used I expect roughly the same outcome from my ARDL approach regarding the direction of the coefficients.

Even though the literature covering meteorological low to medium frequency weather shocks with respect to economic parameters is growing, most of them remain focused on more general economic indicators like GDP, inflation and overall price levels. To my knowledge, only two papers are using the theory of storage framework, and none of them has included exogenous shocks like the ENSO-cycle until now. According to Cashin et al. (2015) the key challenge in studying a climate-economy relationship is "identification", i.e. distinguishing the effects of climate on economic activity from many other factors potentially co-varying with it. This can be one reason why little literature regarding this specific phenomenon in relation to commodity prices is available. In addition, the widely-spread global cultivation of wheat can cause the more locally induced impacts of ENSO to remain unnoticed as the effects differ per region and time-frame. Moreover, possible interrelations can either strengthen certain effects or compensate them to the extent of offsetting or even reversing them, making it hard to distinguish individual relationships. This is confirmed by Abdolrahimi (2016) as she concludes that ENSO-shocks are largely heterogeneous and its impacts vary substantially depending on phase and crop type. Nevertheless, based on the combination of sound economic concepts regarding the supply-and-demand dynamics and the above-mentioned results regarding crop production, inflation and GDP growth in the U.S. with respect to El Niño shocks, I expect that the individual warm and cold phases of the ENSO-cycle will yield negative and positive effects respectively.

Although El Niño and La Niña are scientifically seen as each other's opposite phase within the ENSO-cycle, more recently evidence is growing that due to the asymmetric manifestations of ENSO both over time and per region, this implies that El Niño and La Niña do not necessarily have to offset each-other's impacts on agricultural, financial and or economic parameters as argued by Mason and Goddard in 2001 and Iizumi et al. in 2014. It is important to take this into account when interpreting the results.

4 Methodology

The theory of storage from Holbrook Working (1949) and Fama and French (1987) is in my opinion one of the most well-known frameworks within the financial discipline for determining price dynamics of physical assets. Furthermore, the ENSO-cycle is a major climatic event potentially playing a critical but unknown role in terms of risk analysis for the future price setting of wheat cereals. Therefore, I combine the two concepts in this research in the form of an ARDL-model specification without constant and without trend, in which I check if the ENSO-intensity factor adds to the explanatory power of the original base-model and if a significant causal relationship can be distinguished over time with respect to the inter-temporal price relations of wheat and with respect to the net-long position of wheat contracts held.

4.1 The theory of storage – base model

Below formula corresponds to the theory of storage as originally introduced by Holbrook Working in 1949:

$$F_{t,T} = S_t + R_{t,T} + W_{t,T} - C_{t,T} \tag{4}$$

Here, $F_{t,T}$ represents the futures price F on time t for delivery of a commodity at time T, similarly S_t denotes the spot price of the same commodity at time t. Next, $R_{t,T}$ represents the interest rate, $W_{t,T}$ stands for the necessary return for storage and $C_{t,T}$ embodies the concept firstly identified as 'convenience stocks' which Working says should explain the continued storage of wheat even at negative returns for storage.

In 1987 Fama and French further build on this theory and rewrite the original equation to test the storage-theory against the expectations-theory, as follows:

$$F_{t,T} - S_t = S_t * R_{t,T} + W_{t,T} - C_{t,T} \tag{5}$$

Hereby creating the currently known 'commodity base' denoted by $F_{t,T} - S_t$ which is also used in the more recent literature of Perales (2010) and Ubilava (2017). This formula assumes the commodity base (i.e. the difference between the futures and spot price) to be equal to the sum of the interest forgone, for the sake of continuity denoted by $S_t * R_{t,T}$ and the marginal warehousing costs denoted by $W_{t,T}$, minus the marginal convenience yield denoted by $C_{t,T}$. Fama and French explain the convenience yield arises because having inventory of a physical commodity like wheat rather than the derivate product like a futures contract can have either productive value, since wheat is used as input for the production of other consumption goods, or can be used to meet unexpected demand. Though, one important implication to keep in mind according to Fama and French is that while controlling for variation in marginal warehousing costs and marginal convenience yields, the maturity period $(T - t)$ basis for any stored commodity should vary on a one-to-one basis with the maturity period $(T - t)$ of the interest rate.

4.2 Unit root tests

To be able to transform the above base-model correctly and successfully into the long-term representation of the ARDL (p, q) estimation model, I first need to check if none of the time-series variables are stationary in the order of I(2) as this would mean the variables cannot be analysed using an ARDL-model specification. To validate that all variables are suited for an ARDL-based regression analysis I perform the standard Augmented Dickey-Fuller (ADF)-test to verify that all variables are either I(0) or I(1).[21] See below hypotheses for the ADF test on stationarity:

$H_{0,\,ADF}$: y ~ I(1), meaning the time series has unit root (random walk), it is non-stationary
$H_{1,\,ADF}$: y ~ I(0), meaning the time series has no unit root, it is stationary

For the ADF-test the null-hypothesis of non-stationarity is rejected at level when the absolute value of the T-statistic is greater than the critical value of the confidence interval against which it is tested, vice versa if the absolute value of the T-statistic is lower, the null-hypothesis cannot be rejected. After establishing that none of the time-series variables is stationary at I(2) I can proceed to specify the ARDL-model and the optimal number of lags for the final regressions.

4.3 The price effect of ENSO

Several papers from Working (1949), Fama and French (1987) and Perales (2010) have provided extensive evidence of the significance of storage costs and interest rates in the relation between spot prices and futures prices of commodities. In addition, a growing collection of more meteorologically oriented literature is distinguishing the magnitude and directional effects of the various ENSO phases on commodity prices, production and yields through various different models. The number of different models and approaches used to identify a significant (causal) relationship between ENSO and these economic parameters indirectly confirms the social search to fully grasp the weather-imposed manifestations and evolution over time of the various phases of the ENSO-cycle within the boundaries of statistical models. The social and economic interest in acquiring information on future major (temporal) climate changes well in advance to prepare for the possible consequences is significant (Yeh et al., 2009).

Therefore, to provide the first corner stones in answering the research question I first investigate the long-run relations between ENSO and the commodity base through hypothesis testing of the following two sub-questions:

1. *Does El Niño have a statistically significant effect on the commodity base?*
   $H_{0,1}: \beta_9 = 0$, El Niño has no statistically significant effect on CB_t
   $H_{\alpha,1}: \beta_9 \neq 0$, El Niño has a statistically significant effect on CB_t

[21] See the work of Dickey and Fuller (1979) for a more mathematical elaboration on the characteristics of the unit root test.

2. *Does La Niña have a statistically significant effect on the commodity base?*

   $\mathrm{H}_{0,2}$: $\beta_{10} = 0$, La Niña has no statistically significant effect on $\mathrm{CB_t}$

   $\mathrm{H}_{\alpha,2}$: $\beta_{10} \neq 0$, La Niña has a statistically significant effect on $\mathrm{CB_t}$

To test the joint significance of the coefficients for both these representations of ENSO simultaneously I use an ARDL specification with no constant and no trend in an effort to approximate the original theory of storage equation from Fama and French as closely as possible:

$$CB_t = \beta_0 CB_{t-1} + \beta_1 I_{t-i} + \beta_2 W_{t-i} + \beta_3 ONI_{t-i} + \beta_4 IN_{t-i} + \beta_5 GDP_{t-i} + \beta_6 dum2008_t + \beta_7 dumELNINO_t + \beta_8 dumLANINA_t + \beta_9 ONI_{t-i} dumELNINO_{t-i} + \beta_{10} ONI_{t-i} dumLANINA_{t-i} + \varepsilon_t \tag{6}$$

Here $CB_t = F_{t,T} - S_t$ and $I_{t,T} = S_t * R_{t,T}$ for the sake of simplicity and understanding.

The wheat commodity base CB_t at time t represents the difference between the spot price at time t for delivery in the same month and the futures price for the same product at time t for delivery at time T in the next month. This is assumed to be equal to the sum of the interest forgone in storing wheat for a one month period denoted by $I_{t,T}$, plus the marginal storage costs for one month denoted by $W_{t,T}$, the ONI-standardised sea surface temperatures (SST's) measured in the i number of months preceding time t represented by ONI_{t-i}, the phase in which the ENSO-cycle was at that same time of measurement through the dummies $dumELNINO_t$ and $dumLANINA_t$ which take each the value of one for El Niño and La Niña phases respectively and zero for all other moments in time, the interaction-effect between the ENSO-phase and the value of the ONI-standardised SST's at that same time of measurement. All corrected for exogenous shocks due to changes in inflation denoted by IN_{t-i} and nominal GDP growth denoted by GDP_{t-i}. In addition, I account for a break in the data due to the financial crisis during 2007-2009 by adding a recession dummy variable $dum2008_{t-i}$ which takes the value of one during recession months and zero otherwise.

In this research I deliberately choose for an approach without constant or intercept term for two reasons. First of all, the theory of storage explicitly states that the commodity base should be equal to warehousing costs, the interest forgone in storing and the convenience yield. Adding a constant term would deteriorate these direct relations between aforementioned factors while yielding no additional explanatory power and practically equal results. In addition, the aim is to mirror the original equation of Fama and French (1987) as best as possible, only adding the ENSO- and control-variables to investigate the added value and explanatory power of this expanded model.

Having established non-stationarity at I(2) for abovementioned time-series variables I will execute an ARDL-model without constant and trend to run the regression analysis using the HAC Newey-West covariance matrix specification to simultaneously correct for possible auto-correlation of the residuals and heteroscedasticity of the residuals.[22] Also, after selection through running an unrestricted Vector Auto-Regressive (VAR) pre-test, I let the Akaike Information Criterion (AIC) automatically determine the optimal lag structure of the model within the chosen boundaries.

When using a monthly frequented dataset, it is often advised to use either six, twelve or twenty-four lags when estimating an ARDL-model. For these specific sub-questions, I choose to use a maximum number of six lags for the regressors for the following reasons. The most common type of wheat produced for which I have obtained the prices is a winter wheat with a cultivation period of seven to eight months, if I include seven or more lags in the model, the influence of the ENSO-intensity on the commodity base may become clouded due to possible deviating price trends for the previous harvest(s) of (other types of) wheat. Furthermore, due to the transition from old harvest stocks to new harvest supply, the level of wheat held on storage can experience a significant shift in a short time which can also further bias the effect of ENSO on prices. For the dependent variable, I use a maximum of two lags.[23]

Subsequently, I conduct the F-Bounds Test on joint significance of the coefficients of the lagged levels of the independent variables as developed by Pesaran, Shin and Smith (2001) to test for the existence of cointegration among the variables, with the following hypotheses:

H_0: $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$, no cointegration of variables

H_1: $\beta_0 \neq \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4 \neq \beta_5 \neq \beta_6 \neq \beta_7 \neq \beta_8 \neq \beta_9 \neq \beta_{10} \neq 0$, cointegration of variables in the long-run

Other than with most statistical tests, it is not the probability of the F-test that is the key to the conclusion of the test. The Bounds Test has a non-standard distribution depending on the order of stationarity of the modelled variables included in the ARDL-model [i.e. I(0) or I(1)], on the number of regressors and on the sample size used. Consequently, the retrieved F-statistic is compared to its corresponding critical values in the form of upper and lower bounds as explained by Pesaran, Shin and Smith (2001). The lower critical bound assumes all variables are I(0) meaning no cointegration between the tested variables exists. The upper bound assumes all variables are I(1) meaning cointegration among the variables is proven. It follows that, if the F-statistic is below the lower bound critical value, the null-hypothesis of no cointegration cannot be rejected. When the outcome of the F-statistic is greater than the upper bound critical value, the null-hypothesis of no cointegration is rejected and the variables are cointegrated. Either of these outcomes would already provide an answer for the abovementioned first two sub-questions. However, a third result is also possible when the value of the obtained F-statistic falls between the lower and upper bound critical values, in this case the test result remains inconclusive.

---

[22] Newey and West (1987) suggest a covariance estimator that is consistent in the presence of heteroscedasticity and autocorrelation (HAC) of unknown form, they assume autocorrelations to die out between distant observations (i.e. if time series are sufficiently lagged the problem of autocorrelation should resolve itself.

[23] The reason for choosing only two lags originates mainly from the fact that the paid student version of EVIEWS 9 SV has a limited memory capacity and computation power, see section on limitations for further explanation.

When cointegration between the variables of the ARDL-regression is found, this justifies a transformation of Equation (6) into an Error-Correction Model (ECM) form. This means transforming the ARDL-model from levels into differences. This way, the newly obtained dependent variable Δy_t corrects for past deviations from the long-run equilibrium relationships. In error correction-form both the short-term and the long-term relationships are directly distinguishable. Moreover, this approach allows for a combination of stationary and non-stationary variables, in contrast to other models like the Johansen cointegration technique. Finally, the ECM-model estimates the error-correction term, also known as the speed of adjustment factor. This term, often denoted as EC_t, embodies the cointegrating relationship between the dependent and independent variables. In addition, the direction of the value for this variable contains information on whether a model is stable or explosive. Consequently, after transformation the error-correction representation of the ARDL-model without constant and trend reads as follows:

$$\begin{aligned}\Delta CB_t = {} & \beta_0 CB_{t-1} + \beta_1 \Delta I_{t-i} + \beta_2 \Delta W_{t-i} + \beta_3 \Delta ONI_{t-i} + \beta_4 \Delta IN_{t-i} + \beta_5 \Delta GDP_{t-i} + \beta_6 dum2008_t + \\ & \beta_7 dumELNINO_t + \beta_8 dumLANINA_t + \beta_9 \Delta ONI_{t-i} dumELNINO_{t-i} + \\ & \beta_{10} \Delta ONI_{t-i} dumLANINA_{t-i} + \lambda EC_{t-1} + \varepsilon_t \end{aligned} \quad \textbf{(7)}$$

Again, also here $CB_t = F_{t,T} - S_t$ and $I_{t,T} = S_t * R_{t,T}$ for the sake of continuity. In addition, the error-correction term EC_{t-1} and its coefficient lambda λ represent the following mathematical relations:

$$EC_{t-1} = CB_{t-i} - \sum_{i=1}^{q} \frac{\beta_i}{\beta_0} X_{i,t} \quad \textbf{(8)}$$

$$\lambda = (1 - \sum_{i=1}^{p} \beta_i) \quad \textbf{(9)}$$

Where $X_{i,t}$ stands for the respective independent variables at time t. The fact that I assume no constant and no trend in this model means the speed of adjustment factor lambda λ is in the end equal to one minus the sum of coefficients of the lagged values of the dependent variable. The delta Δ represents the first differences value of each variable, in short this means for the dependent variable that $\Delta CB_t = CB_t - CB_{t-1}$. Furthermore, coefficients β_1 through β_{10} represent the short-term dynamic coefficients of the model's long-run equilibrium. The error-correction term EC_{t-1} exists of the extracted residuals from the aforementioned long-run ARDL Equation (6) and therefore forms the cointegration factor or long-term representation of the error-correction model. The error-term's coefficient λ is the model's 'speed of adjustment' factor which should be negative. This is a prerequisite for the model to return to a long-term equilibrium after short-term deviations in its parameters. Apart from that, all mentioned variables represent the same underlying values as in Equation (6). Hence, the wheat commodity base CB_t at time t represents the difference between the spot price at time t for delivery in the same month and the futures price for the same product at time t for delivery at time T in the next month. This is assumed to be equal to the sum of the interest forgone in storing wheat for a one month period denoted by $I_{t,T}$, plus the marginal storage costs for one month denoted by

$W_{t,T}$, the ONI-standardised sea surface temperatures (SST's) measured in the i number of months preceding time t represented by ONI_{t-i}, the phase in which the ENSO-cycle was at that same time of measurement through the dummies $dumELNINO_t$ and $dumLANINA_t$ which take each the value of one for El Niño and La Niña phases respectively and zero for all other moments in time, the interaction-effect between the ENSO-phase and the value of the ONI-standardised SST's at that same time of measurement. All corrected for exogenous shocks due to changes in inflation denoted by IN_{t-i} and nominal GDP growth denoted by GDP_{t-i}. Moreover, I account for a break in the data due to the financial crisis during 2007-2009 by adding a recession dummy variable $dum2008_{t-i}$ which takes the value of one during recession months and zero otherwise.

If the coefficient (λ) of the error correction term EC_{t-1} is negative and significant, it means the model readjusts to its mean and is therefore stable. If it is positive, the model is explosive and implies that the model is not stable. I use the Cumulative Sum (CUSUM) and the Cumulative Sum Squared (CUSUMSQ) tests from Brown, Durbin and Evans (1975) which provide a plot of the cumulative sum of (squared) recursive residuals in relation to the 5% confidence interval from the zero line to verify if the model remains within its confidence boundaries meaning that it is stable and reverts to a long-term equilibrium after short-term shocks or deviations in its parameters.

Finally, the explanatory powers of the ARDL and ECM models and the model's stability validations of this research' price effect-model are compared to the respective outcomes of the theory of storage base model to further substantiate the added value and statistical significance of the ENSO parameters.

4.4 The net long position effect of ENSO

As shortly mentioned in section 2.2.2, the total open interest of all non-commercial traders is generally understood as a reliable indicator of market sentiment, because the majority of these traders is actually interested in taking delivery of the underlying assets of the futures contracts held in position. The only reservation on this parameter is that open interest is foremost an indicator of cash in- and outflows of the futures market which analysts use to confirm the strength of a certain trend at that moment (CME Group, 2020), but this indicator does not provide for a direction of this possible trend. To gain a more accurate understanding of the direction of a possibly present trend I focus on the difference between the total long contracts held as opposed to the total short contracts held by the non-commercial traders, known as a net long position. If this difference is positive (i.e. the majority of non-commercial market participants is holding long contracts) it is generally assumed that, prices are expected to increase in the future and or the availability of the underlying asset may become scarcer in the near future, which ceteris paribus ultimately also leads to an increase in prices. In contrast, when the difference is negative (i.e. the majority of market participants is holding short contracts) it is generally assumed that, prices are expected to decrease in the future and or the availability of the underlying asset may become abundant in the near future, which ceteris paribus ultimately also leads to a decrease in prices. For this reason, the net long position of all futures contracts held by non-commercial traders at a certain point in time provides for a forward-looking indicator of price trends. This informative and predictive value has already been proven in previous literature from

Hong and Yogo (2012) amongst others, hence based on this evidence I utilise the net long position of non-commercial open interest to determine the indirect effects of ENSO on the wheat price dynamics.

Combined with the meteorological capability to forecast major climate events up until six months in the future as declared by Ludescher et al. (2013) amongst others, identifying a significant relationship between the phase and intensity of the ENSO-cycle and the net long position of non-commercial open interest would also imply an effect on wheat prices. Therefore, subjecting the following two additional sub-questions to hypothesis-testing yields the other part of the foundation on which I can draw the final conclusion for the main problem statement of this research.

3. *Does El Niño have a statistically significant effect on the non-commercial net long position?*
   $H_{0,3}$: $\beta_9 = 0$, El Niño has no statistically significant effect on $NCNL_t$
   $H_{\alpha,3}$: $\beta_9 \neq 0$, El Niño has a statistically significant effect on $NCNL_t$

4. *Does La Niña have a statistically significant effect on the non-commercial net long position?*
   $H_{0,4}$: $\beta_{10} = 0$, La Niña has no statistically significant effect on $NCNL_t$
   $H_{\alpha,4}$: $\beta_{10} \neq 0$, La Niña has a statistically significant effect on $NCNL_t$

Here I use the same ARDL regression equation without constant and trend to test the joint significance of El Niño's and La Niña's coefficients, only in relation to a different dependent variable:

$$NCNL_t = \beta_0 NCNL_{t-1} + \beta_1 I_{t-i} + \beta_2 W_{t-i} + \beta_3 ONI_{t-i} + \beta_4 IN_{t-i} + \beta_5 GDP_{t-i} + \beta_6 dum2008_t + \beta_7 dumELNINO_t + \beta_8 dumLANINA_t + \beta_9 ONI_{t-i} dumELNINO_{t-i} + \beta_{10} ONI_{t-i} dumLANINA_{t-i} + \varepsilon_t \qquad \textbf{(10)}$$

Where $I_{t,T} = S_t * R_{t,T}$ continues to represent the interest forgone in storing.

The abbreviation $NCNL_t$ stands for the Non-Commercial Net Long position of long and short contracts held and not yet off-set by delivery or (re)sale prior to delivery, understood as the number of long contracts at time t minus the number of short contracts held by non-commercial market participants at time t. As in the previous regression, the net long position is assumed to be equal to the sum of the interest forgone in storing wheat for a one month period denoted by $I_{t,T}$, plus the marginal storage costs for one month denoted by $W_{t,T}$, the ONI-standardised sea surface temperatures (SST's) measured in the i number of months preceding time t represented by ONI_{t-i}, the phase in which the ENSO-cycle was at that same time of measurement through the dummies $dumELNINO_t$ and $dumLANINA_t$ which take each the value of one for El Niño and La Niña phases respectively and zero for all other moments in time, the interaction-effect between the ENSO-phase and the value of the ONI-standardised SST's at that same time of measurement. All corrected for exogenous shocks due to changes in inflation denoted by IN_{t-i} and nominal GDP growth denoted by GDP_{t-i}. In addition, I account for a break in the data due to the financial crisis during 2007-2009 by adding a recession dummy variable $dum2008_{t-i}$ which takes the value of one during recession months and zero otherwise. Again I make no use of a constant term or intercept because I aim to expand on the original theory of storage equation from Fama and French (1987) and compare these results to the base model, in addition the differences in results are so small that they are negligible.

Likewise, I use a maximum number of two lags for the dependent variable and six lags for the regressors, partly for the same reasons, but partly also because at the time of writing the general consensus in meteorological literature states that major climate events like ENSO cannot be reliably forecasted more than six months ahead. Hence, possible effects of ENSO which lie more than six months ahead cannot reasonably be expected to be taken into consideration by the market participants at the time of their decision-making process. After selection through running an unrestricted Vector Auto-Regressive (VAR) pre-test, I let the Akaike Information Criterion (AIC) automatically determine the optimal lag structure of the model within the chosen boundaries.

Similarly, having established non-stationarity at I(2) for abovementioned time-series variables, I repeat the same ARDL-model regression without constant and trend and same F-Bounds Test as explained in section 4.3 from Pesaran, Shin and Smith (2001) to test for cointegration with monthly frequented data. When the F-statistic is smaller than the lower bound critical value, the null-hypothesis of no cointegration cannot be rejected. When the F-statistic is greater than the upper limit critical value, cointegration is proven and I can proceed to rewrite into the ECM-representation to test further for long- and short-term dynamic relationships and the speed of adjustment factor as also previously explained.

If all coefficients of the independent variables are not statistically significant different from zero, the null-hypothesis of no cointegration cannot be rejected and implies no significant causal relationship between ENSO and the wheat price commodity base. In contrast, if all coefficients are statistically significant different from zero, cointegration among the variables exists. This provides the base to transform the equations into an ECM representation to test for short-term and long-term dynamic relationships and model stability, as follows:

$$\begin{aligned}\Delta NCNL_t = {} & \beta_0 NCNL_{t-1} + \beta_1 \Delta I_{t-i} + \beta_2 \Delta W_{t-i} + \beta_3 \Delta ONI_{t-i} + \beta_4 \Delta IN_{t-i} + \beta_5 \Delta GDP_{t-i} + \\ & \beta_6 dum2008_t + \beta_7 dumELNINO_t + \beta_8 dumLANINA_t + \beta_9 \Delta ONI_{t-i} dumELNINO_{t-i} + \\ & \beta_{10} \Delta ONI_{t-i} dumLANINA_{t-i} + \lambda EC_{t-1} + \varepsilon_t \end{aligned} \qquad \textbf{(11)}$$

Also here $I_{t,T} = S_t * R_{t,T}$ still represents the interest forgone in storing. In addition, the error-correction term EC_{t-1} and its coefficient lambda λ represent the following mathematical relations:

$$EC_{t-1} = NCNL_{t-i} - \sum_{i=1}^{q} \frac{\beta_i}{\beta_0} X_{i,t} \qquad \textbf{(12)}$$

$$\lambda = (1 - \sum_{i=1}^{p} \beta_i) \qquad \textbf{(13)}$$

Where $X_{i,t}$ stands for the respective independent variables at time t. Given I assume no constant and no trend in this model the speed of adjustment factor lambda λ is in the end equal to one minus the sum of coefficients of the lagged values of the dependent variable. The delta Δ represents the first differences value of each variable, in short this means for the dependent variable that $\Delta NCNL_t = NCNL_t - NCNL_{t-1}$. Furthermore, coefficients β_1 through β_{10} represent the short-run dynamic coefficients of the model's long-run equilibrium. The error-correction term EC_{t-1} exists of the extracted residuals from the aforementioned long-run ARDL Equation (10) and therefore forms the cointegration factor or long-run representation of the error-correction model. The error-term's coefficient λ is the model's 'speed of adjustment' factor which should be negative. This is a prerequisite for the model to return to a long-term equilibrium after short-term deviations in its parameters. Apart from that, all mentioned variables represent the same underlying values as in Equation (10). Hence, the term $NCNL_t$ represents the Non-Commercial Net Long position of long and short contracts held and not yet off-set by delivery or (re)sale prior to delivery, understood as the number of long contracts at time t minus the number of short contracts held by non-commercial market participants at time t. This is assumed to be equal to the sum of the interest forgone in storing wheat for a one month period denoted by $I_{t,T}$, plus the marginal storage costs for one month denoted by $W_{t,T}$, the ONI-standardised sea surface temperatures (SST's) measured in the i number of months preceding time t represented by ONI_{t-i}, the phase in which the ENSO-cycle was at that same time of measurement through the dummies $dumELNINO_t$ and $dumLANINA_t$ which take each the value of one for El Niño and La Niña phases respectively and zero for all other moments in time, the interaction-effect between the ENSO-phase and the value of the ONI-standardised SST's at that same time of measurement. All corrected for exogenous shocks due to changes in inflation denoted by IN_{t-i} and nominal GDP growth denoted by GDP_{t-i}. Moreover, I account for a break in the data due to the financial crisis during 2007-2009 by adding a recession dummy variable $dum2008_{t-i}$ which takes the value of one during recession months and zero otherwise.

If the coefficient (λ) of the error correction term EC_{t-1} is negative and significant, it means the model readjusts to its mean and is therefore stable. If it is positive, the model is explosive and implies that the model is not stable. Similar to the price-effect model, I use the Cumulative Sum (CUSUM) and the Cumulative Sum Squared (CUSUMSQ) tests from Brown, Durbin and Evans (1975) which provide a plot of the cumulative sum of (squared) recursive residuals in relation to the 5% confidence interval from the zero line to verify if the model remains within its confidence boundaries meaning that it is stable and reverts to a long-term equilibrium after short-term shocks or deviations in its parameters.

4.5 Robustness checks

To further substantiate the added value and statistical significance the results from the price-effect model, I perform two robustness checks by repeating the price-effect model analysis using the SOI index and TNI index respectively as independent variable to measure the intensity of the ENSO phases. This way I can compare the models' explanatory power, stability and significance of the variables and the error-correction term with those of the price-effect model based on the ONI measurement for ENSO.

Finally, I run an ARDL model for the base model from section 4.1 where Equation (5) serves as the regression input. I use the same data for the commodity base, interest forgone and marginal warehousing costs and the used model specifications remain equal. Through comparison of the price-effect model with the base model I can establish if the explanatory power and significance of variables has increased through adding the ENSO and control variables as compared to the recreated version of the original theory of storage.

Based on the combination of the results for all four hypotheses, their respective model stability validations and the robustness checks I can formulate the final answer to the central problem statement of this research. But first, I further elaborate on the choice of data and variables for this research in the next section.

5 Data

For this research, I use two separate but mostly overlapping datasets. One set is used to test the hypotheses regarding the effects of ENSO on the wheat price commodity base and the second dataset is used to test for the hypotheses regarding the effects of ENSO on the non-commercial net long position. As such, the only two differences between the two datasets concern the dependent variable and the period of testing. The crucial reason why there is a difference between the period of testing originates in the fact that somehow the CBOT started documenting the futures prices for wheat only in 2006, hence to obtain the commodity base as per the used definition I can only start modelling the regression from 2006 onwards. Opposed to the futures price data, all other variables know a much longer history. Hence, in an effort to derive the best and most accurate possible outcome of the regression analysis by testing the influence of ENSO on the non-commercial net long position, the second dataset dates back to 1995. This also helps to include the often cited very strong El Niño period of 1997-98 within the test period.

5.1 Variables and controls

It follows from the above that for all four hypotheses of this research the same independent variables, control variables and dummies are used. All data regarding the spot price, futures price, non-commercial open interest, interest rate, storage rate, nominal GDP, inflation and the ENSO-intensity indices used in this research originate from and are applicable to the United States of America, therefore providing for a coherent and solid foundation to draw conclusions on. For all variables in this research, I use a monthly frequency of data points. Below I elaborate in more detail on each of these variables used in this research.

5.1.1 Wheat futures and spot prices

For the first two hypotheses regarding the ENSO effect on prices I use a combination of data for two variables, the spot prices and futures prices for the exact same commodity: wheat. For the spot price, I use the United States Department of Agriculture (USDA) U.S. wheat No.2 Soft Red Winter (SRW) prices, denoted in U.S. Dollars per bushel, from April 1982 until April 2020 with a monthly frequency. This type of wheat is the most widely cultivated throughout the U.S. and because its cultivation period of roughly seven to eight months as compared to four months for other wheat types it is much more suitable for a study involving lagged regression analysis. As explained in section 2.2.1, wheat spot prices are comparable to forward prices for delivery in the same month, inherently this means spot prices have a one-month maturity.

For the futures price, I use the officially coupled derivative called the Wheat Continuous Average price of the same product, denoted in U.S. Dollar cents per bushel, from April 2006 to April 2020, provided by the e-CBOT. This is the electronic trading platform created and operated by the CBOT. As for the spot price, also the futures prices have a one-month maturity and are retrieved with a monthly frequency. For the sake of simplicity, I transform the unit notation of the futures prices time series also to U.S. Dollars per bushel to match with the spot prices.

I specifically choose for U.S. wheat spot and futures prices foremost because I need a set of a physical price and derived futures price that are accurately linked with each other and with the other relevant data that is essential in this research to conduct a reliable empirical study. Furthermore, to arrive at the dependent variable of this study, I subtract the spot price value from the futures price value of the same month to obtain the difference between the two variables, generally known as the ‘commodity base’.

5.1.2 Non-commercial net long position

For the remaining two hypotheses regarding the ENSO effect on the non-commercial net long interest I use data from the CFTC’s COT-report on open interest of U.S. Wheat. In particular, I use the total long and total short number of contracts held by non-commercial market participants as dependent variable, denoted in number of contracts of five thousand bushels per contract. From the beginning this information was firstly documented in April 1995 to April 2020. Given the COT-report on open interest is published on a weekly basis, the monthly frequented data retrieved also represents a monthly average like the spot and futures prices. As the number of contracts held covers all positions worldwide of this particular commodity the numbers are extremely high, hence to mitigate the magnitude of this variable and prevent for a biased regression I divide the number of contracts held by ten thousand. Therefore, one unit of non-commercial long interest reads as ten thousand contracts of five thousand bushels. Finally, subtracting the total short open interest from the total long open interest yields the net long position of contracts held by non-commercial market participants at any point in time, which makes for the alternative commodity base as dependent variable for the second part of this empirical study. As explained in section 2.2.2 of this paper the net long position of non-commercial market participants in particular provides for a reliable and generally accepted predictor for the actual asset prices. As such, finding a statistically significant (dynamic) relationship between ENSO and the non-commercial net long position would indirectly also prove or reconfirm a causal relationship between ENSO and the price dynamics of wheat.

However, as Hong and Yogo (2012) also point out, the open interest data has a pro-cyclical character and is correlated with macroeconomic activity. Therefore, to correct for these interlinkages and obtain a more accurate estimation of the ENSO-effect, I add the nominal GDP growth and rate of inflation to the regression equations as control variables for the business cyclicality and macroeconomic activity.

5.1.3 Interest rate

The interest rate, also known as the time value of money, plays a significant role both in the theory of storage as in the actual forward and futures contracts of physical commodities. While the interest rate is more often used as a control variable in regressions, in this research based on the theory of storage it is explicitly modelled as an essential part of the relation between futures and spot prices. Although the international commodity trading industry uses the 3-Month United States Dollar LIBOR + 200 basis points as the industry standard for interest rate in calculating prices for wheat contracts (CME Group, 2009), I deliberately choose to work with the United States Dollar 1-Month Deposit Rate, because the theory of storage requires that the maturity of the interest rate equals that of the other parameters in the equation. The data for the interest rate is retrieved from the Thomson Reuters database with monthly frequency, ranging from January 1975 to April 2020 and denoted as a percentage rate.

5.1.4 Marginal warehousing costs

The data for the actual marginal storage costs is obtained through direct contact with the CFTC, where Dr. Kunda of the Agricultural section of the Markets Intelligence Branch (MIB) has provided the actual historic data of the marginal storage fees for U.S. Wheat No.2 SRW in U.S. Dollar cents per bushel per month, ranging from pre-1989 to April 2020. This data is retrieved in monthly frequency and transformed from cents to dollars per bushel per month to match with the futures and spot prices. As one of the three main pillars of the theory of storage this data is crucial for the quality and reliability of this research. In line with expectations a clear break is visible in the data before and after the year 2010 when the CME Group introduced the VSR system for storage rates to improve the convergence between spot and futures prices.

5.1.5 Marginal convenience yield

For the marginal convenience yield no data exists. The marginal convenience yield of storage, as explained by Fama and French (1987), represents a theoretical concept implying that stored wheat contains productive value or can be used to meet unexpected demand. This implied value depends on numerous factors and can change on a daily basis due to any fluctuation in these parameters. As such, it is impossible to assign values to this conceptual part of the equation on beforehand. It is probably best compared to the concept of technological innovation, which is also assumed to be existent but not possible to explicitly model as a variable in a regression. Whereas the technological innovation is often assumed to be implied by the real GDP growth, the marginal convenience yield of storage can possibly best be assumed and approximated by the error-correction term of the ECM-model used to test all four hypotheses in this research, for two reasons. Firstly, it ensures the Martingale Restriction is met as it makes the equation hold at all times. Secondly, the error-correction term is always assumed to have a negative coefficient as this implies a stable model which is mean-reverting. Likewise, the convenience yield is also generally assumed to have a negative relation to the commodity base as Working (1949) and Fama and French (1987) also state.

5.1.6 Inflation

The inflation rate is obtained through transforming the Consumer Price Index (CPI), which is a weighted average index of price changes for a fixed basket of consumer goods and services, like transport, food and healthcare. The data on CPI for this study is retrieved from the United States Bureau of Economic Analysis (BEA) in monthly frequency, ranging from January 1913 to February 2020. More specifically, I use the CPI for all urban consumers (CPI-U), for all items, weighed per the U.S. city average, not seasonally adjusted. This index is averaged with respect to the 1982-84 base period. The positive (negative) percentage change in CPI between one period and another is called inflation (deflation), shown below:

$$Inflation = \frac{(CPI_t - CPI_{t-1})}{CPI_{t-1}} \times 100\% \qquad \textbf{(16)}$$

Where t and $t-1$ denote the months over which the inflation (deflation) is calculated. Inflation (deflation) implies also the rate at which the purchasing power of the domestic currency is falling (rising).

5.1.7 Nominal GDP growth

Economic growth reflects a cyclical but consistent increase in the capacity of an economy to produce goods and services, compared from one period in time to another. The productive capacity of a nation is called a nation's Gross Domestic Product (GDP), which can be defined and measured in three different ways: according to the expenditure approach, the production approach or the income approach (Burda and Wyplosz, 2009). The most commonly used method is the expenditure approach, which entails the sum of consumption, private domestic investments, government expenditures and net exports (i.e. total exports minus total imports). This approach produces the well-known concept of nominal GDP (i.e. not adjusted for inflation) that I use in this study. The data obtained for the nominal GDP is provided by IHS Markit, an independent and leading online macroeconomic information database with access to the U.S. BEA archive to correctly convert the original quarterly GDP data into an index of monthly averages without adjusting for seasonality. The latter is particularly useful for this research. Hence, the data is monthly frequented and ranges from January 1992 to February 2020. Furthermore, I have transformed this data according to the below formula (17) to arrive at the growth rate as used for the regression analyses of this study:

$$Nominal\ GDP\ growth\ = \frac{(GDP_t - GDP_{t-1})}{GDP_{t-1}} \times 100\% \tag{17}$$

Where t and $t-1$ denote the months over which the GDP growth is calculated. The nominal GDP growth rate provides a useful insight into the general direction and magnitude of growth for the overall economy, not adjusted for inflation. The nominal GDP growth rate and rate of inflation I calculate separately on purpose here to better identify the controlling value of each of these two parameters in the regression equation and to improve the reliability of the outcome for the coefficients and significance of the independent variables in this research.

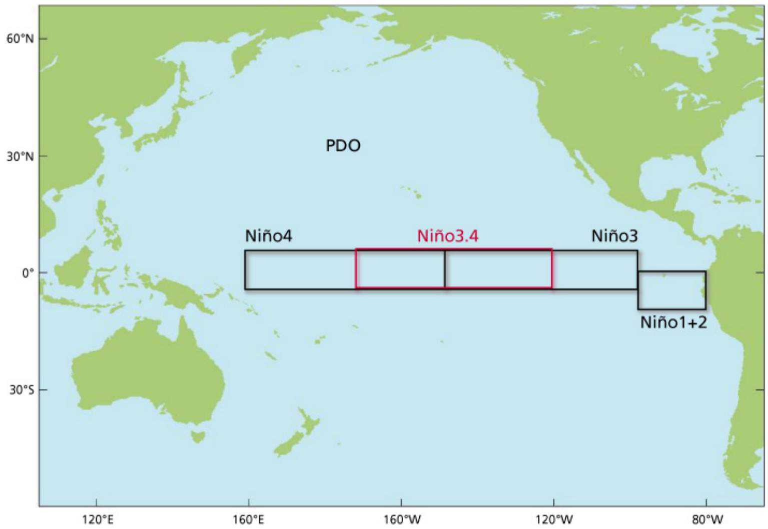
5.2 The ENSO-intensity variables

Whereas the concept is relatively new in the field of financial-economic science, the ENSO-cycle has been studied already for decades from the meteorological and climatological point of view as discussed in the theory section. The result yields the three main ways to measure the duration and intensity of (extreme) ENSO- phases listed earlier in section 2.1.5. I prefer to use the Oceanic Niño Index (ONI) as the ENSO-intensity variable over the SOI and TNI indices, which I motivate in detail below in the next sub-section. The other two indices describing ENSO (SOI and TNI) will be used to run robustness checks.

5.2.1 Oceanic Niño Index (ONI)

As per the officially used definition of the NOAA Climate Prediction Center (CPC), this index identifies an El Niño phase as characterised by a minimum of five consecutive 3-month running mean of Sea Surface Temperature (SST) anomalies in the Niño-3.4 region (see Fig. 2) that is greater or equal to the threshold of +0.5°C and lower or equal to -0.5°C for a La Niña phase, with respect to a centered 30-year base period updated every 5 years (NOAA CPC, 2020). This means the average reading of each month is averaged with the readings of the preceding and following month, subsequently this value is compared to the 30-year base reference average to arrive at the index value. When this index value exceeds the threshold criteria for five months in a row this indicates a major ENSO event. When the index value lies within the [-0.5°C, +0.5°C] threshold criteria, neutral conditions are assumed.

Figure 2: Regions of the Pacific Ocean monitored by the NOAA for SST anomalies.



As one of the world's second most important source of weather variation according to Glantz (2001), Goddard and Dilley (2005) and others, the ENSO-cycle is studied extensively already for many years and a common practice has emerged that the Niño-3.4 region provides the most accurate information to predict possible effects of ENSO-events.[24]

Thus, as opposed to both older and more recent literature from Wyrtki (1975), Brunner (2002), Cashin et al. (2015) and Bekkering (2017) which all use the SOI to measure ENSO activities and in contrast to the cause-effect relationship between the atmospheric and oceanic conditions as explained in the theory section of this paper, the official definition of ENSO-phases is based on the SST's and not on the atmospheric Southern-Oscillation component. Consequently, to safeguard the accuracy of the data and reliability of this empirical study I follow the official NOAA's practices. Furthermore, a recent study from Ubilava (2017) in which also the ENSO effects on asymmetries in wheat price dynamics are researched also uses the ONI index for the regression analysis. The ONI data obtained ranges from January 1870 until January 2020 with a monthly frequency and are logically denoted as index values.

5.2.2 Southern Oscillation Index (SOI)

Also in line with the NOAA, for the robustness check in this research the SOI is explained as the normalised pressure difference between Tahiti, French Polynesia and Darwin, Australia. Though several variations of calculated SOI values exist from various climate centres, here I use the SOI from the Climatic Research Unit (CRU) data that follows NOAA's standards of calculation, based on the method given by Ropelewski

[24] For a detailed explanation on the different regions distinguished in the equatorial Pacific Ocean, see the International Research Institute for Climate and Society's web page on ENSO Essentials (IRI, 2020).

and Jones (1987). Anomalies are considered as departures from the 1951-1980 base period. With that in mind, the following equations show how the SOI-values are established according to the NOAA:

$$SOI = \frac{Standardised\ Tahiti - Standardised\ Darwin}{MSD}$$

Where its components are derived as follows:

$$Std.Tahiti = \frac{\chi_{Tahiti\ SLP} - \mu_{Tahiti\ SLP}}{SD_{Tahiti}}$$

$$SD\ Tahiti = \sqrt{\Sigma(\chi_{\text{Tahiti SLP}} - \mu_{Tahiti\ SLP})^2/N}$$

$$Std.Darwin = \frac{\chi_{Darwin\ SLP} - \mu_{Darwin\ SLP}}{SD_{Darwin}}$$

$$SD\ Darwin = \sqrt{\Sigma(\chi_{\text{Darwin SLP}} - \mu_{Darwin\ SLP})^2/N}$$

$$MSD = Monthly\ Standard\ Deviation = \sqrt{\Sigma(\text{Standardised Tahiti} - Standardised\ Darwin)^2/N^*}$$

The above formulas are provided by the NOAA (2017).

Here, N stands for the number of months, N^* of the last equation represents the number of summed months and SLP for the Sea Level (Air) Pressure values (in hPa) at Tahiti and Darwin respectively. The data used is derived from the CRU with a monthly frequency ranging from January 1866 to February 2020. In contrast to ONI, negative values of SOI represent El Niño periods and positive values represent La Niña periods.

5.2.3 Trans-Niño Index (TNI)

The TNI is defined as the difference in normalised SST anomalies between the Niño-1+2 region and the Niño-4 region, with a five-month running mean applied as opposed to the three-month running mean of the ONI. This five-month running mean is subsequently standardised using the 1950-1979 period. As touched upon in section 2.1.5, Trenberth and Stepaniak (2001) argue that the Niño 3.4 index should be used in combination with the TNI index they introduced, as this would improve the ability to explain for the evolution of El Niño phases in particular. However, the literature does not seem to have picked up on this suggestion as the vast majority still uses either the ONI or SOI indices for their research. The data obtained for the TNI values is also of monthly frequency, ranges from January 1870 to November 2019 and is also derived from NOAA's online database.

5.3 Dummy variables

To correct for exogenous shocks in the data and explicitly model implied information otherwise remaining invisible I create a few essential dummies to improve the quality and reliability of the regression analyses.

U.S. financial crisis dummy

Given both datasets cover the period during which the well-known U.S. 2007-2009 financial crisis occurred and all data is derived from and applicable to the U.S. market, it is very likely that one or more time series of variables will show a significant structural break around the 2007-2009 period. To correct for this possible break, it is statistically desirable to add a dummy variable which identifies the 2007-2009 recession period and corrects for the break caused in the data resulting from this temporal shock.

The National Bureau of Economic Research (NBER) monitors key economic indicators to trace the state of the business cycle. The real GDP growth rate is considered as the primary indicator, though also factors like employment, interest rates and household income are taken into account. According to the NBER, economy contractions start at the peak and end at the bottom of the business cycle (NBER, 2020). However, a contraction is only labelled a recession if the contraction period is longer than twelve months. Consequently, for the regression analyses in this paper with data ranging from 1995 and 2006 respectively until 2020, this means only the December 2007 until June 2009 period is marked as an economic recession. Therefore, a dummy variable is constructed using the NBER definition as a basis, which takes the value of one during this particular recession period and zero for the months preceding and following this period.

El Niño dummy

The ENSO-cycle phenomenon with its three very differently characterised ocean-atmosphere coupled phases provides for a complex variable to model in a linear regression model. Hence, to distinguish from their respective differing impacts and to prevent the mitigation of possible significant individual effects due to counter-effects caused by the previous and or following phase, it is crucial to explicitly identify and model the warm and cold phases of the ENSO-cycle. Therefore, in line with the NOAA and their official calculation method for the ONI index as used in this research, I create an El Niño dummy that has a value of one when the five consecutive 3-month running mean of Sea Surface Temperature (SST) anomalies in the Niño-3.4 region is greater or equal to the threshold of +0.5°C with respect to the 30-year base period and a value of zero otherwise.

La Niña dummy

Likewise, I do the same for the cold phase of the ENSO-cycle. In line with NOAA's official calculation method for the ONI index as used in this research, I create a La Niña dummy that has a value of one when the five consecutive 3-month running mean of Sea Surface Temperature (SST) anomalies in the Niño-3.4 region is lower or equal to -0.5°C with respect to the 30-year base period and a value of zero otherwise.

It follows then, that when the ONI index values lie in between the two threshold criteria, both dummies take the value of zero as neutral weather conditions are assumed. Furthermore, the ENSO-phase dummies derived from the ONI calculation technique will also be applied when running robustness checks with the SOI and TNI indices in order to maintain equal periods in time where El Niño and La Niña phases are occurring and hence also identified as such.

5.4 Final datasets

Following the explanations of all the independent, control and dummy variables, see below in Table 1 the final dataset used for the ARDL and ECM models ran to test the hypotheses formulated to answer sub-questions 1 and 2. All data for both datasets consists of continuous time series with a monthly frequency.

Table 1: Summary of variables for the price-effect model.

Summary of variables used for the regression analyses of ENSO-impacts on the wheat price commodity base of U.S. No.2 SRW Wheat. Including information on the variables and symbols used in this paper, the type of variable and the unit measurement of each variable. The test period for this dataset ranges from April 2006 to November 2019 at monthly frequency.

Variable	Symbol	Type	Unit
Commodity base	CB	Dependent	US$/bu
Interest forgone	I	Independent	US$/bu
Storage cost	W	Independent	US$/bu/month
ONI index	ONI	Independent	Index
SOI index	SOI	Independent	Index
TNI index	TNI	Independent	Index
Inflation rate	IN	Control	%
Nom. GDP growth	GDP	Control	%
Recession dummy	dum2008	Dummy	Binary
El Niño dummy	dumELNINO	Dummy	Binary
La Niña dummy	dumLANINA	Dummy	Binary

Likewise, Table 2 shows the final dataset used for the ARDL and ECM models ran to test the hypotheses formulated to answer sub-questions 3 and 4 respectively.

Table 2: Summary of variables for the net long position-effect model.

Summary of variables used for the regression analyses of ENSO-impacts on the net long position of non-commercial market participants in the futures and spot market for U.S. No.2 SRW Wheat. Including information on the variables and symbols used in this paper, the type of variable and the unit measurement of each variable. The test period for this dataset ranges from April 1995 to November 2019 at monthly frequency. The notation "bu" is the industry abbreviation for "bushel" of wheat grains.

Variable	Symbol	Type	Unit
Non-commercial net long position	NCNL	Dependent	x10.000 contracts of 5000 bu
Interest forgone	I	Independent	US$/bu
Storage cost	W	Independent	US$/bu/month
ONI index	ONI	Independent	Index
SOI index	SOI	Independent	Index
TNI index	TNI	Independent	Index
Inflation rate	IN	Control	%
Nom. GDP growth	GDP	Control	%
Recession dummy	dum2008	Dummy	Binary
El Niño dummy	dumELNINO	Dummy	Binary
La Niña dummy	dumLANINA	Dummy	Binary

5.5 Descriptive statistics

Before running regression analyses, it is always essential to first obtain an informative overview of the basic characteristics of each variable. Table 3 below therefore provides the descriptive statistics for the price-effect model where the wheat price commodity base is the dependent variable for the regression analyses:

Table 3: Descriptive statistics of the price-effect model dataset.

The descriptive statistics for the common sample of the price-effect model data for the test period of April 2006 to November 2019. A skewness value lower than zero means the time series is negatively skewed, a value of zero represents a perfect normal distribution, a value greater than zero means the time series is positively skewed. A kurtosis value lower than 3 means the time series is platykurtic, a kurtosis value of 3 represents normal kurtosis implying a perfect normal distribution, a value greater than 3 means the time series is leptokurtic. The Jarque-Bera test tests for the combination of skewness and kurtosis with a null-hypothesis assuming the data is normally distributed and an alternative hypothesis assuming no normal distribution.

Variables	Mean	Median	Max.	Min.	Std. Dev.	Skewness	Kurtosis	Jarque-Bera	Prob.	Sum	Sum Sq. Dev.	Obs.
CB	0.774	0.610	3.203	-1.656	0.782	0.936	4.774	45.440	0.000	126.889	99.715	164
I	0.077	0.025	0.465	0.005	0.097	1.842	6.084	157.755	0.000	12.592	1.518	164
W	0.078	0.050	0.200	0.045	0.042	1.411	4.321	66.314	0.000	12.737	0.281	164
ONI	0.023	-0.085	2.570	-1.790	0.883	0.460	3.453	7.183	0.028	3.760	127.174	164
IN	0.155	0.169	1.008	-1.915	0.383	-1.113	7.866	195.692	0.000	25.396	23.867	164
GDP	0.282	0.298	1.770	-1.731	0.529	-0.232	4.267	12.444	0.002	46.326	45.647	164
dum2008	0.116	0.000	1.000	0.000	0.321	2.401	6.763	254.253	0.000	19.000	16.799	164
dumELNINO	0.262	0.000	1.000	0.000	0.441	1.081	2.169	36.677	0.000	43.000	31.726	164
dumLANINA	0.299	0.000	1.000	0.000	0.459	0.879	1.773	31.417	0.000	49.000	34.360	164
ONI_ELNINO	0.296	0.000	2.570	0.000	0.586	2.170	7.246	251.888	0.000	48.490	55.932	164
ONI_LANINA	-0.277	0.000	0.000	-1.790	0.484	-1.625	4.517	87.900	0.000	-45.460	38.163	164

It follows quickly that none of the variables are normally distributed as proven by the Jarque-Bera test which tests for skewness and kurtosis combined. A probability value below 0.05 means the null-hypothesis of normal distribution is rejected. As shown, all values are well below this critical value of 0.05. In addition, the individual figures for skewness and kurtosis confirm the same. Although GDP and ONI skewness values are close to zero, the requirement for normal distribution, all other variables are clearly different from zero. A positive value here means the data of the variable is positively skewed, i.e. the visual representation of the data points will demonstrate a 'long right tail', a negative value means the data of the variable is negatively skewed, i.e. the visual representation of the data points will demonstrate a 'long left tail'. Furthermore, a value of 3 is required according to the kurtosis index for the data points to follow a normal distribution, a value lower than 3 indicates the time series of the variable is platykurtic and a value greater than 3 implies the data follows a leptokurtic distribution. As most of the data is already standardised or differenced no real outliers are detected.

Next, Table 4 below illustrates the same overview for the net long position-effect model which encompasses the non-commercial net long position as dependent variable for the regression analyses. Similar to the previous table, here it is also obvious that none of the variables are normally distributed according to the Jarque-Bera test. However, non-normality does not pose a problem to an ARDL-model nor to its ECM-specification. Even though the non-commercial net long position comes very close to a skewness of zero. Strangely its sum of squared deviations, i.e. the variance, is extremely high. This implies a high volatility in non-commercial open interest. Again, also in this dataset many variables are standardised or differenced and as a result no real outliers can be detected.

Table 4: Descriptive statistics of the net long position-effect model dataset.

The descriptive statistics for the common sample of the net long position-effect model data for the test period of April 1995 to November 2019. A skewness value lower than zero means the time series is negatively skewed, a value of zero represents a perfect normal distribution, a value greater than zero means the time series is positively skewed. A kurtosis value lower than 3 means the time series is platykurtic, a kurtosis value of 3 represents normal kurtosis implying a perfect normal distribution, a value greater than 3 means the time series is leptokurtic. The Jarque-Bera test tests for the combination of skewness and kurtosis with a null-hypothesis assuming the data is normally distributed and an alternative hypothesis assuming no normal distribution.

Variables	Mean	Median	Max.	Min.	Std. Dev.	Skewness	Kurtosis	Jarque-Bera	Prob.	Sum	Sum Sq. Dev.	Obs.
NCNL	-0.610	-0.494	14.767	-14.329	4.296	0.115	4.547	30.183	0.000	-180.596	5443.814	296
I	0.101	0.079	0.465	0.005	0.091	1.114	4.005	73.715	0.000	30.036	2.426	296
W	0.063	0.045	0.200	0.045	0.035	2.272	7.843	543.958	0.000	18.677	0.359	296
ONI	-0.012	-0.105	2.570	-1.790	0.876	0.612	3.559	22.348	0.000	-3.550	226.336	296
IN	0.180	0.186	1.222	-1.915	0.344	-0.899	7.825	326.997	0.000	53.218	34.944	296
GDP	0.360	0.369	1.770	-1.731	0.532	-0.355	4.026	19.221	0.000	106.508	83.504	296
dumELNINO	0.243	0.000	1.000	0.000	0.430	1.197	2.433	74.643	0.000	72.000	54.486	296
dumLANINA	0.318	0.000	1.000	0.000	0.466	0.784	1.614	53.987	0.000	94.000	64.149	296
dum2008	0.064	0.000	1.000	0.000	0.246	3.556	13.648	2022.176	0.000	19.000	17.780	296
ONI_ELNINO	0.280	0.000	2.570	0.000	0.585	2.265	7.499	502.738	0.000	82.980	100.835	296
ONI_LANINA	-0.289	0.000	0.000	-1.790	0.476	-1.482	4.092	123.109	0.000	-85.450	66.903	296

As both tables also show for each dataset at the bottom, I create two interaction terms in which I multiply the dummies for El Niño and La Niña periods with the actual ENSO variable ONI used to measure the intensity of the phase.

6 Results

The results of the regression analyses are structured in the same way as presented in the methodology section. First I present and discuss the unit root test results for the price effects of ENSO, followed by the results regarding the net long position effect of ENSO. Next I outline the results of the ARDL regression analyses, F-Bounds tests and ECM representations for both the price-effect- and net long position effect-models. I also compare the outcomes of both models with those of the theory-of-storage base model. Furthermore, I discuss the model stability validation for both error-correction models on the price effect and net long position effect of ENSO. Finally, I conclude with a robustness check to validate the reliability of the used model-specifications and variables used for the research regressions.

6.1 Unit root tests

Table 5 below shows the results for the ADF-test on stationarity for each individual variable of the price-effect model where I investigate the effects of ENSO on the commodity base. As I use the model specification without constant and without trend, the most important parts of the below table are the ADF-test outputs below the 'None'-indication. Only results that reject the null-hypothesis of non-stationarity at the 5% or 1% significance level are considered reliable and statistically significant to minimise the risk of falsely rejecting or accepting the null-hypothesis. It follows that except for the interest rate forgone in storing I, the marginal storage costs W and the recession dummy and $dum2008$ the ADF-test fails to reject the null-hypothesis of a unit root at level but rejects the null at first differences for all variables. Hence, the interest forgone and the marginal storage costs time series are stationary only at I(1). For all other variables time series I can strongly reject the null-hypothesis at level meaning they are stationary at I(0). Finally, I can conclude that none of the variables are stationary at the order I(2). Therefore, I can proceed to analyse the ARDL-model regression.

Table 5: Unit root test results for the price-effect model.

*ADF-test results per individual variable for the Augmented Dickey-Fuller (ADF) test for the period ranging from April 2006 to November 2019. The lag selection is based on the Akaike Information Criterion (AIC) with a maximum of 6 lags. C denotes constant, C+T denotes constant and (linear) trend, 'None' means no constant and no trend. Superscripts *, ** and *** represent the statistical significance of the ADF t-statistic at the 10%, 5% and 1% significance level, respectively. The null-hypothesis assumes the existence of unit root.*

	Level			First differences		
Variables	**C**	**C + T**	**None**	**C**	**C + T**	**None**
CB	-3.204**	-3.503**	-1.994**	-6.156***	-6.155***	-6.179***
I	-1.814	-1.419	-1.709*	-4.308***	-4.465***	-4.297***
W	-2.695*	-2.639	-1.200	-4.545***	-4.561***	-4.559***
ONI	-3.499***	-4.330***	-3.505***	-7.158***	-7.139***	-7.178***
IN	-7.075***	-7.094***	-6.667***	-7.641***	-7.616***	-7.666***
GDP	-6.261***	-6.446***	-2.423**	-8.147***	-8.121***	-8.174***
dum2008	-2.223	-2.559	-2.092	-12.649***	-12.616***	-12.689***
dumELNINO	-3.336**	-3.421*	-2.676***	-11.930***	-11.885***	-11.963***
dumLANINA	-3.930***	-4.123***	-3.250***	-13.751***	-13.716***	-13.794***
ONI_ELNINO	-3.859***	-3.956**	-2.824***	-9.039***	-9.003***	-9.066***
ONI_LANINA	-3.117**	-3.577**	-2.462**	-7.639***	-7.630***	-7.664***

Likewise, Table 6 provides the outcomes of the ADF-test for each individual variable of the net long position-effect model, which contains the non-commercial net long position as dependent variable. Because this dataset encompasses a different period range in time the characteristics of the variables also show differences. As a result, for a model-specification without constant and trend, here the ADF-test only fails to reject the null-hypothesis at level for the interest rate forgone in storing I and the marginal storage costs W, which are stationary at I(1). For all other variables the ADF-test rejects the null-hypothesis at level at the 1% or 5% significance level and are thus I(0). Again, ADF rejects the null-hypothesis for all variables at first differences which means none of the variables are stationary at I(2). Thus, I can proceed to run the ARDL-model also for this dataset.

Table 6: Unit root test results for the net long position-effect model.

*ADF-test results per individual variable for the Augmented Dickey-Fuller (ADF) test for the period ranging from April 1995 to November 2019. The lag selection is based on the Akaike Information Criterion (AIC) with a maximum of 6 lags. C denotes constant, C+T denotes constant and (linear) trend, 'None' means no constant and no trend. Superscripts *, ** and *** represent the statistical significance of the ADF t-statistic at the 10%, 5% and 1% significance level, respectively. The null-hypothesis of the ADF-test assumes the existence of unit root.*

	Level			**First differences**		
Variables	**C**	**C + T**	**None**	**C**	**C + T**	**None**
NCNL	-4.559***	-5.660***	-4.429***	-9.790***	-9.855***	-9.793***
I	-2.303	-2.165	-1.903*	-5.807***	-5.874***	-5.792***
W	-3.180**	-3.524**	-1.479	-6.175***	-6.170***	-6.185***
ONI	-5.320***	-5.415***	-5.331***	-7.373***	-7.360***	-7.383***
IN	-8.326***	-8.480***	-9.142***	-10.863***	-10.844***	-10.882***
GDP	-4.606***	-8.982***	-2.140**	-11.863***	-11.841***	-11.884***
dum2008	-2.912**	-2.903	-2.819***	-17.088***	-17.060***	-17.117***
dumELNINO	-4.487***	-4.478***	-3.876***	-17.088***	-17.062***	-17.117***
dumLANINA	-4.706***	-4.799***	-3.845***	-9.758***	-9.736***	-9.773***
ONI_ELNINO	-5.470***	-5.458***	-4.615***	-8.665***	-8.653***	-8.680***
ONI_LANINA	-3.925***	-3.992***	-3.228***	-9.697***	-9.677***	-9.713***

6.1 The price effect of ENSO

Table 7 below provides the results of the ARDL model without constant and trend for the price effects of ENSO on the commodity base as explained in the methodology section 4.3. Clearly both El Niño and La Niña have a strong statistically significant impact on the price convergence between the wheat futures and spot prices (i.e. the commodity base). Though many would expect the El Niño and La Niña phases of the ENSO cycle to have opposite effects due to their different climatic characteristics they both manifest negative influences on the commodity base of -0.4246 and -0.5536 respectively, though at relatively different points in time. As this concerns the outcomes of the ARDL regression the coefficients represent the long-term influences of the independent variables on the dependent variable. In addition, the statistical significance, magnitude and direction of long-term coefficients of the other variables are also given in Table 7. According to expectations the interest forgone has a strong positive influence on the commodity base, in contrast the marginal warehousing costs have a negative impact while Fama and French concluded on a positive effect. Nevertheless, since there is no data regarding the marginal convenience yield of storage this implied factor with a claimed negative influence may play a role here. The (adj.) R-squared implies that roughly (88 percent) 90 percent of the dependent variable's variance is explained by the independent variables of the model which is very high. By using the HAC-Newey West covariance matrix, the regression analysis shows very little to no signs of serial correlation within the residuals as the generally accepted rule-of-thumb in statistics reads that a Durbin-Watson statistic between 1.5 and 2.5 implies no serial correlation.

Table 7: ARDL regression analysis output for the price-effect model.

*Long-run ARDL (2-6-3-6-5-0-4-2) regression output for the period April 2006 to November 2019 with the commodity base denoted as CB as dependent variable and ONI as intensity index for the measurement of ENSO activity. A HAC-Newey-West covariance matrix with Bartlett kernel specification is used. Maximum lags for the dependent variable is 2. Automatic optimal lag structure for the regressors is based on the Akaike Information Criterion (AIC) up to a maximum of 6 lags. Here the independent variables I, W, ONI, IN, GDP, ONI_ELNINO and ONI_LANINA are dynamic regressors and the dummies dum2008, dumELNINO and dumLANINA are fixed regressors. Superscripts *, ** and *** represent the statistical significance at the 10%, 5% and 1% significance level, respectively.*

Independent variables	Coefficient	Std. Error	t-Statistic	Regression characteristics	
CB(-2)	0.1229	0.0885	1.3882	R-squared	0.9085
I(-6)	4.1412***	1.5112	2.7405	Adjusted R-squared	0.8802
W(-3)	-5.4983**	2.4602	-2.2349	S.E. of regression	0.2754
ONI(-6)	-0.1581	0.1121	-1.4100	Sum squared resid.	9.1002
IN(-5)	0.1866***	0.0698	2.6715	Durbin-Watson stat.	2.2071
GDP	0.0902**	0.0357	2.5235	Akaike info criterion	0.4646
ONI_ELNINO(-4)	-0.4246***	0.1194	-3.5577	Schwarz criterion	1.2012
ONI_LANINA(-2)	-0.5536***	0.2069	-2.6758	Hannan-Quinn criter.	0.7637
dum2008	0.3189***	0.1014	3.1466		
dumELNINO	0.0215	0.1000	0.2146		
dumLANINA	0.0824	0.0932	0.8845		

To confirm if all independent variables are jointly significant and inherently reconfirm that the coefficients of El Niño and La Niña respectively are indeed significantly different from zero, Table 8 below shows the test results of the Bounds Test for the price-effect model.

Table 8: F-Bounds test and t-Bounds test results for the price-effect model.

Test results of the F-Bounds and t-Bounds tests on cointegration with respect to the ARDL (2-6-3-6-5-0-4-2) model for the price effect of ENSO on the commodity base measured by ONI as intensity index for ENSO. Per significance level ranging from 10% to 1% the lower and upper bounds as explained in Pesaran, Shin and Smith (2001) provide the critical values against which to test the test-statistics. A value below the I(0) lower bound means there is no cointegration and thus the null hypothesis is accepted. A value above the I(1) upper bound means there is cointegration between the variables and hence the null hypothesis is rejected. A value in between the lower and upper bounds renders the test inconclusive. Value of 'k' stands for the number of independent variables in the ARDL regression.

F-Bounds Test

	Value	Signif.	I(0)	I(1)
F-statistic	8.5957	10%	1.70	2.83
k	7	5%	1.97	3.18
		2.5%	2.22	3.49
		1%	2.54	3.91

t-Bounds Test

	Value	Signif.	I(0)	I(1)
t-statistic	-4.4256	10%	-1.62	-3.90
		5%	-1.95	-4.23
		2.5%	-2.24	-4.54
		1%	-2.58	-4.88

The F-test statistic is significantly greater than the I(1) upper bound critical value which forms the threshold to reject the null-hypothesis, I can reject the null-hypothesis of no cointegration even at the 1% significance level. Likewise, the t-test also rejects the null-hypothesis of no cointegration up until the 5% significance level, at confidence intervals of 2.5% and 1% the t-test remains inconclusive. However, this is enough proof to proceed rewriting the ARDL model into an ECM representation, see Table 9 below. The most important result is the direction and significance of the coefficient lambda λ for the long-term cointegration equation-factor EC, also known as the error-correction term, which is significant at the 1% significance level and has a value of -0.1994. This implies that regardless of the impact of short-term shocks, the current model with the used specifications always reverts to a long-term equilibrium value and the magnitude of the coefficient represents the 'speed of adjustment' with which this model returns to that long-term equilibrium state.

Table 9: Error Correction Model representation of the price-effect model.

*Error Correction Model (ECM) of the ARDL (2-6-3-6-5-0-4-2) regression for the period April 2006 to November 2019 with the commodity base denoted as CB as dependent variable and ONI as intensity index for the measurement of ENSO activity. A HAC-Newey-West covariance matrix with Bartlett kernel specification is used. Maximum lags for the dependent variable is 2. Automatic optimal lag structure for the regressors is based on the Akaike Information Criterion (AIC) up to a maximum of 6 lags. Here the independent variables I, W, ONI, IN, ONI_ELNINO and ONI_LANINA are the short-term dynamic regressors and the dummies dum2008, dumELNINO and dumLANINA are fixed regressors. The "Δ" in front of each variable stands for the fact that first differences are taken from each variable and the negative number between parenthesis stands for the number of lags. The error-correction term (EC) forms the long-term representation of the ECM-equation and its coefficient value (λ) represents the speed of adjustment factor. Superscripts *, ** and *** represent the statistical significance at the 10%, 5% and 1% significance level, respectively.*

Independent variables	**Coefficient**	**Std. Error**	**t-Statistic**	**Regression characteristics**	
Δ (CB(-1))	-0.1229*	0.0688	-1.7854	R-squared	0.7083
Δ (I)	-7.9262***	0.9009	-8.7978	Adjusted R-squared	0.6394
Δ (I(-1))	-2.5041**	1.1448	-2.1874	S.E. of regression	0.2677
Δ (I(-2))	-0.6736	0.9731	-0.6922	Sum squared resid.	9.1002
Δ (I(-3))	0.2782	0.9647	0.2884	Durbin-Watson stat.	2.2071
Δ (I(-4))	0.8840	0.9591	0.9217	Akaike info criterion	0.3760
Δ (I(-5))	-4.1412***	0.9252	-4.4762	Schwarz criterion	0.9769
Δ (W)	4.0234*	2.2758	1.7679	Hannan-Quinn criter.	0.6200
Δ (W(-1))	5.0429**	2.2860	2.2060		
Δ (W(-2))	5.4983**	2.3010	2.3896		
Δ (ONI)	-0.0083	0.1398	-0.0593		
Δ (ONI(-1))	-0.4065***	0.1481	-2.7442		
Δ (ONI(-2))	-0.2792**	0.1302	-2.1450		
Δ (ONI(-3))	-0.3818***	0.1280	-2.9821		
Δ (ONI(-4))	0.1487	0.1161	1.2805		
Δ (ONI(-5))	0.1581	0.1038	1.5231		
Δ (IN)	0.1186	0.0745	1.5919		
Δ (IN(-1))	-0.4715***	0.1014	-4.6512		
Δ (IN(-2))	-0.4144***	0.0888	-4.6662		
Δ (IN(-3))	-0.1129	0.0832	-1.3564		
Δ (IN(-4))	-0.1866**	0.0792	-2.3568		
Δ (ONI_ELNINO)	0.0539	0.1441	0.3742		
Δ (ONI_ELNINO(-1))	0.2894*	0.1513	1.9134		
Δ (ONI_ELNINO(-2))	0.4482***	0.1396	3.2099		
Δ (ONI_ELNINO(-3))	0.4247***	0.1347	3.1518		
Δ (ONI_LANINA)	-0.0714	0.1703	-0.4192		
Δ (ONI_LANINA(-1))	0.5536***	0.1703	3.2501		
dum2008	0.3189***	0.1000	3.1887		
dumELNINO	0.0215	0.0501	0.4281		
dumLANINA	0.0824	0.0540	1.5260		
EC(-1)	**-0.1994*** **	**0.0234**	**-8.5309**		

The differenced variables form the short-term part of the equation and represent the short-term shocks or deviations from the equilibrium value of the commodity base. The cointegration equation as represented by EC forms the long-term part of the ECM model and yields the following formula:

$$EC_{t-1} = CB_{t-i} - (-2.0797 * I + 2.4397 * W + 1.7977 * ONI + 3.5077 * IN + 0.3724 * GDP - 1.7213 * ONI\_ELNINO - 0.9329 * ONI\_LANINA)$$

Together with Table 9 this equation shows that in the short-run El Niño and La Niña have a positive deviating impact at different points in time and negative effects in the long-run. Based on the significance of all outcomes, I can reject both null-hypotheses from sub-questions 1 and 2. Hence, El Niño and La Niña have a statistically significant effect on the commodity base CB_t.

This is in line with expectations as external shocks usually lead to short-lived increased insecurity which widens the gap between spot and futures prices. In addition, it takes time for a capital market to incorporate the new information in the pricing-mechanism. In the long-run, however, El Niño and La Niña both have a negative (converging) impact on the commodity base. This can be a direct effect of the climatic conditions caused by ENSO, or as Ludescher et al. claim in 2013, this can be due to the fact that the effects of the recurring phases are accurately predicted and hence anticipated by the market participants.

6.2 The net long position effect of ENSO

Table 10 below provides the results of the ARDL model without constant and trend for the effects of ENSO on the non-commercial net long position as explained in the methodology section 4.4. Again, El Niño and La Niña have a significant negative impact of -2.64 and -1.98 at the 1% and 5% significance level respectively on the dependent variable, the non-commercial net long position in this case. With the non-commercial net long position as dependent variable the negative effects of the hot and cold phases of ENSO are more in line with expectations as changing parameters and increasing insecurity due to extreme weather circumstances would naturally lead to more caution in the decision-making process of market participants in taking any (new) short or long positions.

The lower (adj.) R-squared value of (70 percent) 71 percent in combination with the strong temporarily increased volatility in the dependent variable in the periods 1995-1998 and 2011-2018 may imply that a (lurking) variable or time-fixed effect is missing in this model. This can be investigated further in new empirical research. I use the HAC-Newey West covariance matrix and again the regression analysis shows no signs at all of serial correlation with a DW-statistic of 1.99 which lies also within the 1.5 to 2.5 interval of no serial correlation. However, as opposed to the previous model it is strange that while the non-commercial net long open interest is more often used in literature as successful proxy for prices and market depth, the interest forgone, marginal warehousing costs and inflation are not statistically significant here. Finally, the extremely high sum of squared residuals implies that despite the Newey-West covariance matrix this regression still suffers from heteroscedasticity. Given no presence of outliers, this can relate to the aforementioned periods of temporal increased volatility in the data of the dependent variable.

Table 10: ARDL output for the net long position-effect model.

*Long-run ARDL (2-1-0-1-0-0-0-0) regression output for the period April 1995 to November 2019 with the commodity base denoted as NCNL as dependent variable. A HAC-Newey-West covariance matrix with Bartlett kernel specification is used. Maximum lags for the dependent variable is 2. Automatic optimal lag structure for the regressors is based on the Akaike Information Criterion (AIC) up to a maximum of 6 lags. Here the independent variables I, W, ONI, IN, GDP, ONI_ELNINO and ONI_LANINA are dynamic regressors and the dummies dum2008, dumELNINO and dumLANINA are fixed regressors. Superscripts *, ** and *** represent the statistical significance at the 10%, 5% and 1% significance level, respectively.*

Independent variables	**Coefficient**	**Std. Error**	**t-Statistic**	**Regression characteristics**	
NCNL(-2)	-0.1233*	0.0650	-1.8964	R-squared	0.7120
I(-1)	-10.1287	6.9531	-1.4567	Adjusted R-squared	0.6987
W	-4.6144	3.0446	-1.5156	S.E. of regression	2.3545
ONI(-1)	-1.0171*	0.5382	-1.8898	Sum squared resid.	1552.2890
IN	0.1444	0.3095	0.4664	Durbin-Watson stat.	1.9868
GDP	-0.7097**	0.2761	-2.5704	Akaike info criterion	4.5970
ONI_ELNINO	-2.6387***	0.8136	-3.2432	Schwarz criterion	4.7724
ONI_LANINA	-1.9751**	0.8912	-2.2163	Hannan-Quinn criter.	4.6673
dum2008	-0.3615	0.3458	-1.0455		
dumELNINO	-0.0609	0.5427	-0.1122		
dumLANINA	-0.3534	0.6328	-0.5585		

I perform the F-Bounds and t-Bounds test to check for joint significance of all independent variables and reconfirm if the El Niño and La Niña coefficients are indeed significantly different from zero. Based on the test results as shown in Table 11 below I can reject the null-hypothesis of no cointegration for both tests at the 1% significance level as the respective test statistics are much greater (in absolute terms) than the upper bound critical values required to reject the null-hypothesis. Despite the lower explanatory power of the ARDL model and as opposed to the lack of statistical significance in many variables, this implies that all independent variables are statistically significant different from zero. This outcome justifies rewriting the ARDL regression for this model also into the ECM representational form.

Table 11: F-Bounds and t-Bounds test results for the net long position-effect model.

Test results of the F-Bounds and t-Bounds tests on cointegration with respect to the ARDL (2-1-0-1-0-0-0-0) model for the net long position effect of ENSO on the non-commercial open interest. Per significance level ranging from 10% to 1% the lower and upper bounds as explained in Pesaran, Shin and Smith (2001) provide the critical values against which to test the test-statistics. A value below the I(0) lower bound means there is no cointegration and thus the null hypothesis is accepted. A value above the I(1) upper bound means there is cointegration between the variables and hence the null hypothesis is rejected. A value in between the lower and upper bounds renders the test inconclusive.

F-Bounds Test

	Value	**Signif.**	**I(0)**	**I(1)**
F-statistic	8.580547	10%	1.70	2.83
k	7	5%	1.97	3.18
		2.5%	2.22	3.49
		1%	2.54	3.91

t-Bounds Test

	Value	**Signif.**	**I(0)**	**I(1)**
t-statistic	-7.5664	10%	-1.62	-3.90
		5%	-1.95	-4.23
		2.5%	-2.24	-4.54
		1%	-2.58	-4.88

The results of the ECM model for the non-commercial net long position effect are shown below in Table 12. Again, the central conclusion can be found at the bottom, here the direction and significance of the coefficient lambda λ for the error-correction term EC is given. Also in this model, the coefficient is negative with a value of -0.2966 and strongly significant at the 1% significance level. This means that despite any short-term shocks, the current model with the used specifications always reverts to a long-term equilibrium value. In addition, the magnitude of the coefficient lambda represents the 'speed of adjustment' with which this model returns to that long-term equilibrium state. Furthermore, the strong drop in explanatory power raises some questions for further investigation.

Table 12: Error Correction Model representation of the net long position-effect model.

*Error Correction Model (ECM) of the ARDL (2-1-0-1-0-0-0-0) regression for the period April 1995 to November 2019 with the commodity base denoted as NCNL as dependent variable. A HAC-Newey-West covariance matrix with Bartlett kernel specification is used. Maximum lags for the dependent variable is 2. Automatic optimal lag structure for the regressors is based on the Akaike Information Criterion (AIC) up to a maximum of 6 lags. The independent variables I and ONI are the short-term dynamic regressors and the dummies dum2008, dumELNINO and dumLANINA are fixed regressors. The "Δ" in front of each variable stands for the fact that first differences are taken from each variable. The error-correction term (EC) forms the long-term representation of the ECM-equation and its coefficient value (λ) represents the speed of adjustment factor. Superscripts *, ** and *** represent the statistical significance at the 10%, 5% and 1% significance level, respectively.*

Independent variables	Coefficient	Std. Error	t-Statistic	Regression characteristics	
Δ (NCNL(-1))	0.1233**	0.0538	2.2940	R-squared	0.2204
Δ (I)	16.8459***	5.9080	2.8514	Adjusted R-squared	0.2041
Δ (ONI)	2.9355***	0.5984	4.9053	S.E. of regression	2.3257
dum2008	-0.3615	0.5748	-0.6289	Sum squared resid.	1552.2890
dumELNINO	-0.0609	0.2767	-0.2200	Durbin-Watson stat.	1.9868
dumLANINA	-0.3534	0.2556	-1.3824	Akaike info criterion	4.5494
EC(-1)	**-0.2966*****	**0.0354**	**-8.3881**	Schwarz criterion	4.6371
				Hannan-Quinn criter.	4.5845

The differenced variables in Table 12 form the short-term part of the equation and represent the short-term shocks or deviations from the equilibrium value of the commodity base. Given the fact that the ONI_ELNINO and ONI_LANINA variables were not lagged according to the automatic lag selection process of the AIC, these variables do not return in the short-term part of the equation as differenced value for the ECM model but remain present in the long-run part of the equation. The cointegration equation as represented by EC forms the long-term part of the ECM model and yields the following formula:

$$EC_{t-1} = NCNL_{t-i} - (22.6494 * I - 15.5588 * W + 6.4685 * ONI + 0.4868 * IN - 2.3929 * GDP \; - 8.8971 * ONI\_ELNINO - 6.6595 * ONI\_LANINA)$$

This equation proves that El Niño and La Niña have negative effects on the non-commercial net long position in the long-run. Based on the significance of all outcomes for this model, I can also reject both null-hypotheses from sub-questions 3 and 4. Hence, El Niño and La Niña have a statistically significant effect on the net long position of non-commercial open interest as denoted by $NCNL$.

As shortly discussed this result is in line with expectations as market participants may want to exercise greater caution when deciding on taking (new) long or short positions. In times of insecurity or external shocks, traders customarily prefer to hedge any current short or long position or wait before making any great changes to their strategies. Hence, the negative effects of El Niño and La Niña suggesting the net long position's value moves back to zero or to its long-term mean matches with these industry habits.

6.3 Model stability validations

Though already partially confirmed by the statistically significant error-correction terms with negative values for their coefficients implying both ECM models are mean-reverting in the long-term, there are two visual tests to reconfirm that both ARDL models with chosen specifications are stable. Through the plots of the Cumulated Sum (CUSUM) and the Cumulative Sum of Squares (CUSUMSQ) Tests on the stability of the coefficients of the estimated ARDL-model, the stability of the model throughout the sample period can be visualised. It follows from the graphs of both models that although the models experience some short-term shocks overall both models are relatively stable, that is their itinerary remains predominantly within the boundaries of the 5% significance level. See Table 13 in Appendix A.

For both the price-effect as the net long position-effect model only a slight deviation of the long-run equilibrium seems visible between roughly 2011 and 2017, which may coincide with the 2012-2013 to 2017-2018 'grain glut'. This was a rather long period of record grain yields and increasing stocks year over year which caused the market prices to drop significantly. Hence, for further investigation and or future research an interesting variable to include in the model would be the local and or global physical stock-levels of wheat, especially given the results of the net long position-effect model. In addition, the difference with respect to the theory of storage base model is also visible in this table.

To further substantiate the results obtained from these models I discuss and compare the outcomes of the robustness checks in the next chapter before concluding on the main research question of this paper.

7 Robustness checks

To verify the robustness of the price-effect model I replace the ONI index variable as intensity measurement for ENSO phases with the SOI index and TNI index respectively to compare the results. In addition, I run the same ARDL model for the base model to compare if the price-effect model as analysed in this research can be seen as an actual improvement or addition to the original theory of storage. Given the questions raised by the valid but debatable outcome of the net long position-effect model there is no use in also doing a robustness check on this model as it is clearly not robust. Nonetheless its significant results do add value to the overall conclusion of this research.

7.1 Price-effect model with SOI index

Tables 14 through 17 in Appendix B show all results for the ARDL model, Bounds tests, ECM and model stability validation graphs from the CUSUM and CUSUMSQ tests for the price-effect model where I use SOI as independent variable. Though the individual variables for El Niño and La Niña do not show any strong signs of significance in the initial ARDL model regression compared to the main regression based on ONI, the Bounds test results still strongly reject the null-hypothesis of no cointegration at the 1% and 2.5% significance level respectively which equals the results shown in Table 8. In addition, the (adj.) explanatory power of the ARDL model based on SOI is still very high at (88%) 90%. Finally, the error-correction term of the rewritten ECM representation retains its negative and strongly significant effect at the 1% significance level and the model retains its stability even though the short-term deviations (see Table 17) are slightly bigger as compared to the same model based on ONI (see Table 13). Therefore, the first check confirms the robustness of the used data and model specifications for the price-effect model.

7.2 Price-effect model with TNI index

Tables 18 through 21 in Appendix C show all results for the ARDL model, Bounds tests, ECM and model stability validation graphs from the CUSUM and CUSUMSQ tests for the price-effect model where I use TNI as independent variable. The individual variables for El Niño and La Niña do not show any strong signs of significance in the initial ARDL model regression, given the lack of significance for El Niño and the very low coefficient value of La Niña. Despite the significance at 5% confidence level, the actual influence of La Niña would be very small to negligible compared to the coefficient value in the price-effect model including ONI as ENSO measure. Nonetheless, the Bounds test results still strongly reject the null-hypothesis of no cointegration at the 1% and 2.5% significance level respectively, resembling the outcomes from Table 8. In addition, the (adj.) explanatory power of the ARDL model is still very high at (87%) 89%. Finally, the error-correction term of the rewritten ECM representation retains its negative and strongly significant effect at the 1% significance level and the model retains its stability even though the short-term deviations (see Table 21) are slightly bigger as compared to the same model based on ONI (see Table 13). Therefore, also the second check confirms the robustness of the used data and model specifications for the price-effect model.

7.3 Comparison to base model

At last I regress the theory of storage base model using Equation (5) as regression equation for the ARDL-model with the same data as from the price-effect model for the dependent variable, the interest forgone in storing and the warehousing costs, uncontrolled for the recession time-fixed effects, the inflation rates and nominal GDP to best mirror the original equation of the theory of storage by Fama and French (1987). This way I can compare the explanatory power, significance of variables and outcome of the ECM representation with the price-effect model based on ONI as measurement for ENSO intensity.

Tables 22 through 25 in Appendix D provide the results obtained from the ARDL model, Bounds tests, ECM and model stability validation graphs from the CUSUM and CUSUMSQ tests for the uncontrolled base model of the theory of storage. It is clear that the (adj.) explanatory power of this model is lower at (85%) 86%, hence the addition of the ENSO variable, ENSO interaction variables, control variables and dummies has a significantly positive effect on the explanatory power meaning the added variables are not redundant. The marginal warehousing costs do show a strongly significant positive influence in the base model as opposed to the result of the price-effect model in Table 7 and as Fama and French claim. Despite the omitted variables in this base model and the lower explanatory power, this model strongly rejects the null-hypothesis of no cointegration for both the F-Bounds and t-Bounds test at the 1% significance level. However, the ECM representation in Table 24 shows a much bigger drop in explanatory power. The model stability validation graphs confirm the lack of explanatory power as they show a much less stable path over time with bigger deviations from the long-term equilibrium. Finally, the error-correction term is negative and strongly significant but the speed of adjustment factor lambda explaining the speed with which this model returns to its long-term equilibrium after short-term shocks is less negative compared to the price-effect model including ENSO. Hence, the information about ENSO and the controlling variables provide for valuable additions to the base model.

8 Conclusion

Based on the combination of former literature outcomes and the results from this empirical research, I draw the final conclusion on the main research question in this chapter. In addition, I explain the limitations this research is subject to and elaborate on suggestions for improvements or further research in the future.

8.1 The final answer

In this research, I investigated the short-run and long-run effects of the El Niño and La Niña phases of the ENSO-cycle on the price dynamics of U.S. wheat and on the market sentiment of wheat as a proxy for the price dynamics of wheat. The problem statement for this research reads as follows: *Do ENSO-cycle weather shocks have a significant impact on the price dynamics of U.S. wheat?* The answer: yes, they do.

This final answer is derived from the results of hypothesis testing four sub-hypotheses divided over two different research models. In the first model called the price-effect model, I investigated the direct impacts of the ENSO-induced weather shocks on the futures-spot price parity of wheat, called the commodity base. The null-hypotheses tested in this model read as follows: i) *El Niño has no significant effect on the commodity base;* and ii) *La Niña has no significant effect on the commodity base*. As the ARDL model coefficients for the El Niño and La Niña variables were significant at the 1% significance level and the Bounds test also strongly rejected the null-hypothesis of no cointegration at the 1% significance level for these ENSO-variables in this model, I rejected both these null-hypotheses of no significant effect for El Niño and La Niña on the commodity base. In other words, based on these results I can confirm that the ENSO-cycle induced extreme weather shocks known as El Niño and La Niña have a significant impact on the price dynamics of wheat both in the short-run and in the long-run. To confirm the stability and reliability of the model and variables used to test these hypotheses, I transformed the ARDL model into an error-correction model which yielded the required negative and statistically significant error-correction term.

In the second model called the net long position-effect model, I researched the effects of ENSO-induced weather shocks on the non-commercial net long position (i.e. net long open interest) as proxy for market sentiment. The null-hypotheses tested in this model read as follows: i) *El Niño has no significant effect on the non-commercial net long position;* and ii) *La Niña has no significant effect on the non-commercial net long position*. As the ARDL model coefficients for the El Niño and La Niña variables were significant at the 1% and 5% significance level respectively and the Bounds test also strongly rejected the null-hypothesis of no cointegration at the 1% significance level for these ENSO-variables in this model, I rejected both these null-hypotheses of no significant effect for El Niño and La Niña on the non-commercial net long position. Since market sentiment in the form of open interest is commonly used and understood as a reliable predictor for the actual price movements of the underlying commodity in previous literature, the results of this analysis hence further substantiated the indirect long-term and short-term effects of El Niño and La Niña on the price dynamics of wheat. In addition, the outcome of a negative and statistically significant error-correction term of the error-correction model assures also this models' stability and reliability of results.

Therefore, based on evidence from previous literature and as confirmed by the results of both empirical models used in this paper to test the hypotheses, I can conclude that the El Niño and La Niña phases of the ENSO-cycle have a statistically significant impact on the price dynamics of U.S. wheat.

As opposed to the consensus in many papers stating that El Niño and La Niña would be each other's opposite phase and would therefore also manifest opposing effects on the tested dependent variables, this research showed that both phases either showcase a negative effect or both showcase a positive effect depending on the point in time. The effects of the hot and cold phases of ENSO on both the commodity base and the non-commercial net long position proved negative in the long-run and slightly positive in the short-run. Though for the latter El Niño and La Niña individually did not show any significant short-term effects, the ENSO-cycle as a whole proved to have a strong positive effect on the non-commercial net long position. Moreover, the strongly significant and negative error-correction terms derived from both analyses confirmed the cointegration of the ENSO-cycle variables with the other independent variables within their separate models in explaining the variance of the commodity base and non-commercial net long position over time respectively, both in the short-run and in the long-run. This implies for both models that they revert to their long-term equilibrium value in the long-run despite shocks or deviations in the parameters in the short-run.

I performed several robustness checks to further substantiate the obtained results in this research. The outcomes of which remained strongly significant and for the vast majority even kept the same direction of impacts for each ENSO phase while other measurement indices for ENSO-intensity were used. Finally, compared to the base model, where the ENSO- and control-variables were left out of the equation, the price-effect model also yielded a higher explanatory power for the dependent variable's variance.

8.2 Limitations

The most troublesome limitation for this paper concerns the inability to explore the full regression analysis capacity as intended. In combination with a maximum of six lags for the regressors I can only use a maximum of two lags for the dependent variable to prevent the model from becoming too big to compute for the software's student version available for use from home under these circumstances. Otherwise the software produces an out-of-memory error, hence I restrict the number of lags for the independent variables to two lags for the most optimal model output given the available options. This inability arises due to the COVID-19 pandemic which has led to a long period of social distancing and an explicit stay-at-home order from the Dutch government. As such, I am limited to using a student-version of Eviews 9 to run the regression analyses for this research. Nonetheless, robust and statistically significant results are obtained.

In comparison to other papers, this research does not break down results to a country- or region-specific level and inherently the possible positive or negative spill-over effects from neighbouring economies of a specific nation or region are not visible nor taken into account. Crop production and or yield from neighbouring countries may be hit less hard or harder than in a given country. Dependent on the trade exposure, the effects of ENSO shocks on net long position and price in one country can either be mitigated or amplified by those neighbouring economies. In addition, one could argue that a better data string to use for the interest rate would be the beginning-of-month yields on U.S. Treasury bills, as used by Fama and French in 1987, to better mirror their original regression of the theory of storage.

From a meteorological point of view, the global climate and weather variations are tremendously complicated and interrelated. Due to the atmospheric global teleconnections between different oscillations, internal positive and negative feedback processes between oceanic and atmospheric conditions could occur. These processes can either reinforce or weaken each other and a wide array of other individual phenomena. Take for example the oceanic Rossby- and Kelvin-waves and their propagation speeds in the Pacific Ocean, which are also said to play an important role in the ENSO-cycle as they represent the ocean's 'adiabatic' response to atmospheric changes as a sum of free and forced waves (Cane, 1983). Such specific aspects are too small and too many to all account or correct for in this research. A very thorough and holistic approach would be needed to include all these kinds of variables, which in turn would probably result in an analysis so complex that it would lack a clear oversight. As Hallegatte et al. (2007) put it, a part of the problem in assessing the impact of climatic events on the economy is due to the fact that quantification of the impacts is still in its infancy. In addition, a new discussion emerged regarding the question if a clear distinction can be made between the 'canonical' El Niño and the recently claimed discovery of El Niño 'Modoki'. Given their believed dynamic differences in spatial and temporal characteristics as in teleconnection patterns, it is difficult to say whether the found impact finds its causation in one of the two or in a combination of both of the climatic events (Ashok et al., 2007; Weng et al., 2007; Ashok and Yamagata, 2009).

On a more environmental note, a fierce point of discussion in most recent years is the question whether or not we, as inhabitants of the Earth, contribute significantly to the extent to which major climate events are said to happen more frequently and in more extreme forms in recent years. If one wants to explore this line of reasoning, the average increase/decrease in climate temperature could be added to the equation.

Finally, from a practical market point of view, I shortly touched upon the fact that the transport and storage market for wheat encompasses a whole supply-and-demand mechanism of its own. As such, not only the marginal storage costs but also the marginal transport costs should be included in the difference between the spot and futures price of wheat. In addition, some literature also suggests the actual storage levels of wheat which represent the magnitude of surplus crops available in the market plays a major role in the price mechanism and is in turn linked to the cultivation cycles of wheat.

8.3 Suggestions

In further research, it would be of additional value to reconfirm if the non-commercial net long position indeed yields a statistically significant causal relationship with the actual spot and futures prices of wheat over time. However, this would require a much larger dataset with many more control and dummy variables to control and correct for all kinds of seasonalities, multiple breaks in the dataset because of economic recessions, changing stock levels, changing storage costs and changing purchasing power both from the importing as from the exporting countries of wheat in order to distil a pure result. In addition, the individual regressions of the four hypotheses in this research can be combined in a new research in which two different models can be tested, one regressing the distilled effect of the non-commercial long position on the commodity base of wheat without ENSO-intensity variables and one model including the ENSO-variables to investigate if the available forecasted information on ENSO-cycle at time $t-6$ is already incorporated in the decision-making process of the market participants or not.

Another essential aspect that can potentially be a limitation for which this research is also partly to blame, is the problem in science that almost every research has the main objective to prove something new instead of repeating formerly conducted research with the idea of checking whether assumptions made and methodologies used especially over 20-40 years ago are still valid. Therefore, I strongly suggest this research to be repeated with the same variables after a few years to see if relations have changed and or if the use of a different regression model can yield better results.

REFERENCES

Abdolrahimi, M. (2016, May). *The Effect of El Niño Southern Oscillation (ENSO) on World Cereal Production.* Retrieved June 28, 2017, from Sydney Digital Theses: https://ses.library.usyd.edu.au//bitstream/2123/15498/1/Abdolrahimi-ma-thesis.pdf

Ashok, K. and Yamagata, T. (2009). The El Niño with a difference. *Nature, 461*, 481-484.

Ashok, K., Behera, S. K., Rao, S. A., Weng, H. and Yamagata, T. (2007, November 8). El Niño Modoki and its possible teleconnection. *Journal of Geophysical Research, 112*(11).

Badjeck, M.-C., Allison, E. H., Halls, A. S. and Dulvy, N. K. (2010, May). Impacts of climate variability and change on fishery-based livelihoods. *Marine Policy, 34*(3), 375-383.

Bekkering, R. (2017, December 22). *The impact of El Niño on stock returns and trading volume of developing countries.* Retrieved from Erasmus University Thesis Repository: hdl.handle.net/2105/41493

Bjerknes, J. (1969, March 1). Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather Review, 97*(3), 163-172.

Breeden, D. T. (1980, May). Consumption Risk in Futures Markets. *The Journal of Finance, 35*(2), 503-520.

Brennan, M. J. (1958, March). The Supply of Storage. *The American Economic Review, 48*(1), 50-72.

Broadie, M., Glasserman, P. and Kou, S. G. (1999, January). Connecting discrete and continuous path-dependent options. *Finance and Stochastics, 3*(1), 55-82.

Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for Testing the Constancy of Regression Relationships over Time. *Journal of the Royal Statistical Society. Series B (Methodological), 37*(2), 149-192.

Brunner, A. D. (2002, February). El Niño and World Primary Commodity Prices: Warm Water or Hot Air? *The Review of Economics and Statistics, 84*(1), 176-183.

Burda, M. and Wyplosz, C. (2009). *Macroeconomics: A European Text* (5th Edition ed.). Oxford: Oxford University Press.

Cai, W., Borlace, S., Lengaigne, M., van Rensch, P., Collins, M., Vecchi, G., . . . Jin, F.-F. (2014, January 19). Increasing frequency of extreme El Niño events due to greenhouse warming. *Nature Climate Change, 4*(2), 111-116.

Cane, M. A. (1983, December 16). Oceanographic Events During El Niño. *Science, 222*(4629), 1189-1195.

Cashin, P., Mohaddes, K. and Raissi, M. (2015, May 1). *Fair Weather or Foul? The Macroeconomic Effects of El Niño.* Retrieved November 18, 2016, from Federal Reserve Bank of Dallas Working Papers: http://www.dallasfed.org/assets/documents/institute/wpapers/2015/0239.pdf

Changnon, S. A. (1999, April 8). Impacts of 1997-98 El Niño Generated Weather in the United States. *Bulletin of the American Meteorological Society, 80*(9), 1819-1827.

CME CBOT. (2017). *Chapter 14: Wheat Futures.* Retrieved Juky 6, 2017, from CME CBOT Rulebook: https://www.cmegroup.com/rulebook/CBOT/II/14/14.pdf

CME Group. (2009, November 30). An Introduction to Variable Storage Rates in Wheat Futures. *Research & Product Development*, 1-4.

CME Group. (2020, April 20). *Introduction to Futures: Open Interest*. Retrieved April 20, 2020, from CME Education: https://www.cmegroup.com/education/courses/introduction-to-futures/open-interest.html

Cootner, P. H. (1960, August). Returns to Speculators: Telser versus Keynes. *Journal of Political Economy, 68*(4), 396-404.

Curtis, B. C., Rajaram, S. and Macpherson, H. G. (2002). *Bread Wheat: Improvement and Production.* Rome: Food and Agriculture Organisation of the United Nations (FAO).

Dusak, K. (1973, December). Futures Trading and Investor Returns: An Investigation of Commodity Market Risk Premiums. *Journal of Political Economy, 81*(6), 1387-1406.

Dybvig, P. H. and Ross, S. A. (1989). Arbitrage. In J. Eatwell, M. Milgate and P. Newman (Eds.), *Finance* (pp. 57-71). Palgrave Macmillan UK.

Fama, E. F. (1970, May). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25*(2), 383-417.

Fama, E. F. and French, K. R. (1987, January). Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage. *The Journal of Business, 60*(1), 55-73.

FAOSTAT. (2017). *Food Balance Sheets: Wheat and other products*. Retrieved August 6, 2017, from FAOSTAT: http://www.fao.org/faostat/en/#data/FBS

Fedorov, A. V. and Philander, S. G. (2000, June 16). Is El Niño Changing? *Science, 288*, 1997-2002.

Food and Agriculture Organisation of the United States. (2017). *Production and Trade statistics*. (Food and Agriculture Organisation of the United States) Retrieved Augsut 6, 2017, from FAO.org - Statistics: http://www.fao.org/economic/ess/ess-trade/en/

Glantz, M. H. (1996). *Currents of Change: Impacts of El Niño and La Niña on Climate and Society.* Cambridge: Cambridge University Press.

Glantz, M. H. (2001). *Currents of Change: Impacts of El Niño and La Niña on Climate and Society* (Second Edition ed.). Cambridge University Press.

Goddard, L. and Dilley, M. (2005, March 1). El Niño: Catastrophe or Opportunity. *Journal of Climate, 18*, 651-665.

Gu, G., Adler, R. F. and Huffman, G. J. (2015, May 7). Long-term changes/trends in surface temperature and precipitation during the satellite era (1979–2012). *Climate Dynamics, 46*, 1091-1105.

Hallegatte, S., Hourcade, J.-C. and Dumas, P. (2007). Why economic dynamics matter in assessing climate change damages: Illustration on extreme events. *Ecological Economics, 62*(2), 330-340.

Hazuka, T. B. (1984, July). Consumption Betas and Backwardation in Commodity Markets. *The Journal of Finance, 39*(3), 647-655.

Holton, J. R. and Dmowska, R. (1989). *El Niño, La Niña and the Southern Oscillation.* (S. G. Philander, Ed.) San Diego: Academic Press.

Hong, H. and Yogo, M. (2012). What does futures market interest tell us about the macoeconomy and asset prices? *Journal of Financial Economics, 105*, 473-490.

Hu, X., Yang, S. and Cai, M. (2016, January 9). Contrasting the eastern Pacific El Niño and the central Pacific El Niño: process-based feedback attribution. *Climate Dynamics, 47*, 2413–2424.

Huisman, R. (2014, January). Finance 2: Financial instruments and strategies to manage risk. Rotterdam.

Hull, J. C. (2013). Chapter 5: Determination of Forward and Futures Prices. In J. C. Hull, *Fundamentals of Futures and Options Markets* (8th Edition ed.).

Iizumi, T., Luo, J.-J., Challinor, A. J., Sakurai, G., Yokozawa, M., Sakuma, H., . . . Yamagata, T. (2014, May 15). Impacts of El Niño Southern Oscillation on the global yields of major crops. *Nature Communications, 5*(3712), 1-7.

IRI. (2020, April 27). *ENSO Essentials*. Retrieved 04 27, 2020, from International Research Institute for Cimate and Society - Earth Institute - Columbia University: https://iri.columbia.edu/our-expertise/climate/enso/enso-essentials/

Julian, P. R. and Chervin, R. M. (1978, October 1). A Study of the Southern Oscillation and Walker Circulation Phenomenon. *Monthly Weather Review, 106*, 1433-1451.

Kaldor, N. (1939, October). Speculation and Economic Stability. *The Review of Economic Studies, 7*(1), 1-27.

Kao, H.-Y. and Yu, J.-Y. (2009, February 1). Contrasting Eastern-Pacific and Central-Pacific Types of ENSO. *Journal of Climate, 22*, 615-632.

Karnauskas, K. B. (2013, October 9). Can we distinguish canonical El Niño from Modoki? *Geophysical Research Letters, 40*, 5246–5251.

Kroner, K. F. and Sultan, J. (1993). Time-Varying Distributions and Dynamic Hedging with Foreign Currency Futures. *Journal of Financial and Quantative Analysis, 8*(4), 535-551.

Kug, J.-S. and Ham, Y.-G. (2011, August 19). Are there two types of La Niña? *Geophysical Research Letters, 38*, 1-6.

Kug, J.-S., Jin, F.-F. and An, S.-I. (2009, March 15). Two Types of El Niño Events: Cold Tongue El Niño and Warm Pool El Niño. *Journal of Climate, 22*, 1499-1515.

Lautier, D. (2008, . .). *The Theory of Storage and the Convenience Yield.* Retrieved July 3, 2019, from Pacific Institute for the Mathematical Sciences: https://www.pims.math.ca/files/Convenienceyield_Lautier.pdf

L'Heureux, M. L., Lee, S. and Lyon, B. (2013, March 10). Recent multidecadal strengthening of the Walker circulation across the tropical Pacific. *Nature Climate Change, 3*, 571–576.

Lobell, M. J. (1942, November). Some observations on The Peruvian coastal current. *Eos, Transactions American Geophysical Union, 23*(2), 332-336.

Longstaff, F. A. (1995). Option Pricing and the Martingal Restriction. *The Review of Financial Studies, 8*(4), 1091-1124.

Ludescher, J., Gozolchiani, A., Bogachev, M. I., Bunde, A., Havlin, S. and Schellnhuber, H. J. (2013, July 13). Improved El Niño forecasting by cooperativity detection. *Proceedings of the national Acadamy of Sciences of the United States of America (PNAS), 110*(29), 11742–11745.

Mason, S. J. and Goddard, L. (2001). Probabilistic Precipitation Anomalies Associated with EN SO. *Bulletin of the American Meteorological Society, 82*(4), 619-638.

Murphy, R. C. (1926, January). Oceanic and Climatic Phenomena along the West Coast of South America during 1925. *Geographical Review, 16*(1).

NAWG. (2017). *Wheat Nutrition*. Retrieved July 2, 2017, from National Association of Wheat Growers: http://www.wheatworld.org/wheat-101/wheat-industry/wheat-nutrition/

Naylor, R., Falcon, W., Wada, N. and Rochberg, D. (2002). Using El Niño-Southern Oscillation climate data to improve food policy planning in Indonesia. *Bulletin of Indonesian Economic Studies, 38*(1), 75-91.

NBER. (2020, April 27). *US Business Cycle Expansions and Contractions*. Retrieved 04 27, 2020, from the National Bureau of Economic Research: https://www-nber-org.eur.idm.oclc.org/cycles.html

Newey, W. K. and West, K. D. (1987). Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review, 28*(3), 777-787.

Ng, V. K. and Pirrong, S. C. (1994). Fundamentals and Volatility: Storage, Spreads, and the Dynamics of Metals Prices. *The Journal of Business, 67*(2), 203-230.

NOAA. (2015, June 17). *How Big is the Pacific Ocean?* Retrieved March 11, 2017, from NOAA Ocean Explorer: http://oceanexplorer.noaa.gov/facts/pacific-size.html

NOAA. (2017, August 17). *Teleconnections*. Retrieved June 28, 2017, from National Oceanic and Atmospheric Administration: https://www.ncdc.noaa.gov/teleconnections/

NOAA CPC. (2020, April 12). *Cold & Warm Episodes by Season*. Retrieved 04 12, 2020, from NOAA National Weather Service - Climate Prediction Center: https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php

Perales, G. B. (2010). The Theory of Storage and Price Dynamics of Agricultural Commodity Futures: the Case of Corn and Wheat. *Ensayos Revista de Economía, 29*(1), 1-22.

Philander, S. G. (1989, October). El Niño and La Niña. *American Scientist, 77*(5), 451-459.

Philander, S. G., Lau, N. C., Pacanowski, R. C., Nath, M. J., Williams, J. B. and Hasselmann, K. (1989, October 27). Two Different Simulations of the Southern Oscillation and El Nino with Coupled Ocean-Atmosphere General Circulation Models [and Discussion]. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 329*(1604), 167-178.

Pidwirny, M. (2006). Introduction to the Atmosphere: El Niño, La Niña and the Southern Oscillation. In M. Pidwirny, *Fundamentals of Physical Geography* (2nd ed.). Retrieved February 3, 2017, from PhysicalGeography.net: http://www.physicalgeography.net/fundamentals/7z.html

Pindyck, R. S. (2001). The Dynamics of Commodity Spot and Futures Markets: A Primer. *The Energy Journal, 22*(3), 1-29.

Podestá, G. P., Messina, C. D., Grondona, M. O. and Magrin, G. O. (1998, October). Associations between Grain Crop Yields in Central-Eastern Argentina and El Niño-Southern Oscillation. *Journal of Applied Meteorology, 38*, 1488-1498.

Rasmusson, E. M. and Wallace, J. M. (1983, December 16). Meteorological Aspects of the El Niño/Southern Oscillation. *Science, 222*(4629), 1195-1202.

Ren, H.-L. and Jin, F.-F. (2011, February 19). Niño indices for two types of ENSO. *Geophyisical Research Letters, 38*, 1-5.

Ren, H.-L., Jin, F.-F., Tian, B. and Scaife, A. A. (2016, October 30). Distinct persistence barriers in two types of ENSO. *Geophysical Research Letters, 43*, 10,973–10,979.

Ropelewski, C. F. and Jones, P. D. (1987, September). An extension of the Tahiti–Darwin southern oscillation index. *Monthly Weather Review, 115*(9), 2161-2165.

Rosenzweig, C., Iglesias, A., Yang, X. B., Epstein, P. R. and Chivian, E. (2000, January). *Climate Change and U.S. Agriculture: The Impacts of Warming and Extreme Weather Events on Productivity, Plant Diseases and Pests.* Retrieved July 2, 2017, from ResearchGate: https://www.researchgate.net/publication/237335094_Climate_Change_and_US_Agriculture_The_Impacts_of_Warming_and_Extreme_Weather_Events_on_Productivity_Plant_Diseases_and_Pests

Rosenzweig, C., Iglesias, A., Yang, X. B., Epstein, P. R. and Chivian, E. (2001, December). Climate change and extreme weather events: Implications for food production, plant diseases, and pests. *Global Change & Human Health, 2*(2), 90-104.

Takahashi, K., Montecinos, A., Goubanova, K. and Dewitte, B. (2011, May 27). ENSO regimes: Reinterpreting the canonical and Modoki El Niño. *Geophysical Research Letters, 38*, 1-5.

Telser, L. G. (1958, June). Futures Trading and the Storage of Cotton and Wheat. *Journal of Political Economy, 66*(3), 233-255.

The Climatic Research Unit. (2017, July 21). *Southern Oscillation Index (SOI).* Retrieved August 15, 2017, from Climatic Research Unit: Data: https://crudata.uea.ac.uk/cru/data/soi/

Trenberth, K. E. (1997, August 1). The Definition of El Niño. *Bulletin of the American Meteorological Society, 78*(12), 2771-2777.

Trenberth, K. E. and Stepaniak, D. P. (2001, April 15). Indices of El Niño Evolution. *Journal of Climate, 14*, 1697-1701.

Ubilava, D. (2017, March 26). The ENSO Effect and Asymmetries in Wheat Price Dynamics. *World Development, 96*, 490-502.

USDA. (2017, July). *Grain: World Markets and Trade.* Retrieved August 4, 2017, from United States Department of Agriculture - Foreign Agricultural Service: https://apps.fas.usda.gov/psdonline/circulars/grain-wheat.pdf

Walker, S. T. and Bliss, E. W. (1932). World Weather. *Memoirs of the Royal Meteorological Society, 4*(36), 53-84.

Wang, C. and Fiedler, P. C. (2006, May 19). ENSO variability and the eastern tropical Pacific: A review. *Progress in Oceanography, 69*, 239–266.

Webster, P. J. and Palmer, T. N. (1997, December 11). The Past and Future of El Niño. *Nature, 390*, 562-564.

Weng, H., Ashok, K., Behera, S. K., Rao, S. A. and Yamagata, T. (2007). Impacts of recent El Nin˜o Modoki on dry/wet conditions in the Pacific rim during boreal summer. *Climate Dynamics, 29*, 113-129.

Working, H. (1949, December). The Theory of Price of Storage. *The American Economic Review, 39*(6), 1254-1262.

Wyrtki, K. (1975, October 1). El Niño - The Dynamic Response of the Equatorial Pacific Ocean to Atmospheric Forcing. *Journal of Physical Oceanography, 5*, 572-584.

Yeh, S.-W., Kug, J.-S., Dewitte, B., Kwon, M.-H., Kirtman, B. P. and Jin, F.-F. (2009, September 24). El Niño in a changing climate. *Nature, 461*, 511-514.

APPENDIX A – Weight conversion and model stability validations

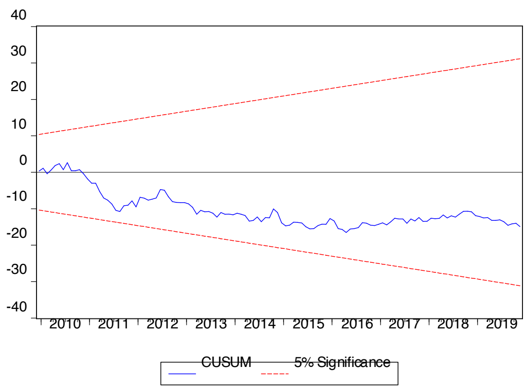
The equivalent of one bushel of wheat which approximately resembles one million wheat kernels is nowadays defined in exact units of weight to be able to calculate in both imperial as metric units.

1 bushel of wheat	= 60 pounds (lbs)	= 0.0272155 metric tons
1 metric ton	= 1000kg	= 36.7437 bushels
1 kg	= 2.20462 lbs	

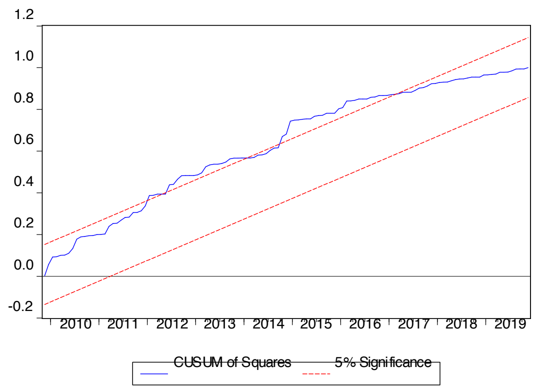
Table 13: Model coefficient stability validations for all models.

ARDL-ECM price-effect model (CB vs ONI)

CUSUM Test

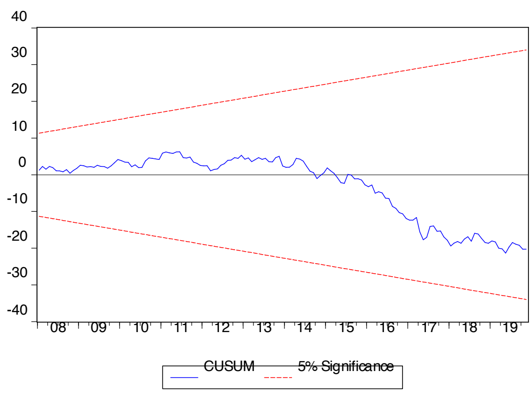


CUSUM of Squares Test

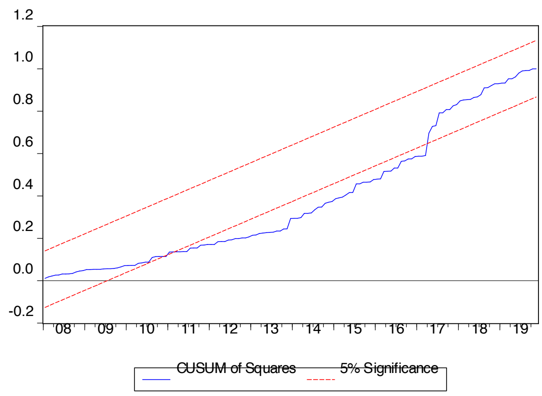


ARDL-ECM non-commercial net long position-effect model (NCNL vs ONI)

CUSUM Test



CUSUM of Squares Test



APPENDIX B – Robustness checks for SOI for the price-effect model

Table 14: ARDL output for price effects of ENSO as measured by the SOI index.

*Long-run ARDL (1-6-3-5-0-2-0-0) regression output for the period April 2006 to November 2019 with the commodity base denoted as CB as dependent variable and SOI as intensity index for the measurement of ENSO activity. A HAC-Newey-West covariance matrix with Bartlett kernel specification is used. Maximum lags for the dependent variable is 2. Automatic optimal lag structure for the regressors is based on the Akaike Information Criterion (AIC) up to a maximum of 6 lags. Here the independent variables I, W, SOI IN, GDP, SOI_ELNINO and SOI_LANINA are dynamic regressors and the dummies dum2008, dumELNINO and dumLANINA are fixed regressors. Superscripts *, ** and *** represent the statistical significance at the 10%, 5% and 1% significance level, respectively.*

Independent variables	Coefficient	Std. Error	t-Statistic	Regression characteristics	
CB(-1)	0.8020***	0.0465	17.2603	R-squared	0.8985
I(-6)	4.4282***	1.4454	3.0635	Adjusted R-squared	0.8784
W(-3)	-4.8516**	2.3260	-2.0858	S.E. of regression	0.2775
IN(-5)	0.1866**	0.0858	2.1758	Sum squared resid.	10.0875
GDP	0.1352***	0.0429	3.1490	Durbin-Watson stat	2.2628
SOI(-2)	-0.0710**	0.0321	-2.2104	Akaike info criterion	0.4284
SOI_ELNINO	0.1270*	0.0747	1.6996	Schwarz criterion	0.9517
SOI_LANINA	-0.0204	0.0652	-0.3135	Hannan-Quinn criter.	0.6409
dum2008	0.3173***	0.1035	3.0655		
dumELNINO	0.1477*	0.0861	1.7166		
dumLANINA	0.0217	0.0672	0.3230		

Table 15: F-Bounds and t-Bounds test results for the price-effect model - SOI index.

Test results of the F-Bounds and t-Bounds tests on cointegration with respect to the ARDL (1-6-3-5-0-2-0-0) model for the price effect of ENSO on the commodity base measured by SOI as intensity index for ENSO. Per significance level ranging from 10% to 1% the lower and upper bounds as explained in Pesaran, Shin and Smith (2001) provide the critical values against which to test the test-statistics. A value below the I(0) lower bound means there is no cointegration and thus the null hypothesis is accepted. A value above the I(1) upper bound means there is cointegration between the variables and hence the null hypothesis is rejected. A value in between the lower and upper bounds renders the test inconclusive.

F-Bounds Test

	Value	Signif.	I(0)	I(1)
F-statistic	8.4249	10%	1.70	2.83
k	7	5%	1.97	3.18
		2.5%	2.22	3.49
		1%	2.54	3.91

t-Bounds Test

	Value	Signif.	I(0)	I(1)
t-statistic	-4.8234	10%	-1.62	-3.90
		5%	-1.95	-4.23
		2.5%	-2.24	-4.54
		1%	-2.58	-4.88

Table 16: Error Correction Model representation of price-effect model - SOI index.

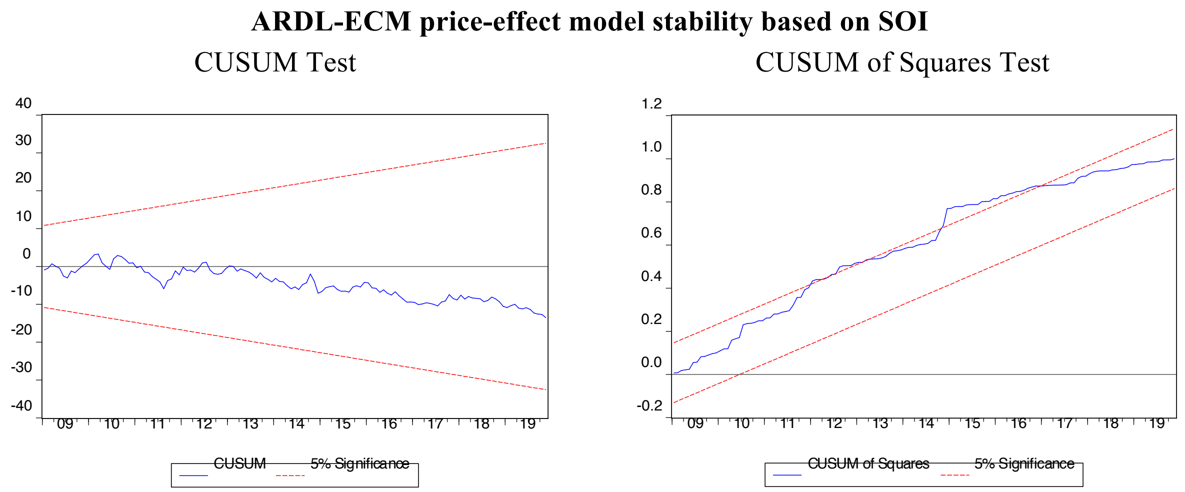
*Error Correction Model (ECM) of the ARDL (1-6-3-5-0-2-0-0) regression for the period April 2006 to November 2019 with the commodity base denoted as CB as dependent variable and SOI as intensity index for the measurement of ENSO activity. A HAC-Newey-West covariance matrix with Bartlett kernel specification is used. Maximum lags for the dependent variable is 2. Automatic optimal lag structure for the regressors is based on the Akaike Information Criterion (AIC) up to a maximum of 6 lags. The independent variables I, W, IN and SOI are the short-term dynamic regressors and the dummies dum2008, dumELNINO and dumLANINA are fixed regressors. The "Δ" in front of each variable stands for the fact that first differences are taken from each variable. The error-correction term (EC) forms the long-term representation of the ECM-equation and its coefficient value (λ) stands for the speed of adjustment to this long-term equilibrium. Superscripts *, ** and *** represent the statistical significance at the 10%, 5% and 1% significance level, respectively.*

Independent variables	**Coefficient**	**Std. Error**	**t-Statistic**	**Regression characteristics**	
Δ (I)	-8.3041***	0.8885	-9.3464	R-squared	0.6767
Δ (I(-1))	-0.9728	0.9885	-0.9841	Adjusted R-squared	0.6322
Δ (I(-2))	-0.1520	0.9547	-0.1592	S.E. of regression	0.2704
Δ (I(-3))	0.2736	0.9408	0.2908	Sum squared resid.	10.0875
Δ (I(-4))	0.5583	0.9303	0.6002	Durbin-Watson stat.	2.2628
Δ (I(-5))	-4.4282***	0.8938	-4.9546	Akaike info criterion	0.3397
Δ (W)	4.9992**	2.2501	2.2218	Schwarz criterion	0.7274
Δ (W(-1))	4.7825**	2.2401	2.1349	Hannan-Quinn criter.	0.4972
Δ (W(-2))	4.8516**	2.2490	2.1572		
Δ (IN)	0.0188	0.0684	0.2754		
Δ (IN(-1))	-0.2454***	0.0820	-2.9936		
Δ (IN(-2))	-0.2937***	0.0772	-3.8025		
Δ (IN(-3))	-0.0678	0.0723	-0.9374		
Δ (IN(-4))	-0.1866***	0.0684	-2.7285		
Δ (SOI)	-0.0792***	0.0263	-3.0095		
Δ (SOI(-1))	0.0711***	0.0263	2.6986		
dum2008	0.3173***	0.0970	3.2729		
dumELNINO	0.1477***	0.0475	3.1125		
dumLANINA	0.0217	0.0480	0.4517		
EC(-1)	**-0.1980*****	**0.0235**	**-8.4262**		

The cointegration equation represented by EC forms the long-run part of the ECM as follows:

$$EC = CB - (-2.9314 * I + 1.7404 * W + 2.5541 * IN + 0.6831 * GDP - 0.4316 * SOI + 0.6414 * SOI\_ELNINO - 0.1032 * SOI\_LANINA)$$

Table 17: CUSUM and CUSUMSQ test results for price-effect model using SOI index.



APPENDIX C – Robustness checks for TNI for the price-effect model

Table 18: ARDL output for the price-effects of ENSO as measured by the TNI index.

*Long-run ARDL (2-6-3-6-5-0-4-2) regression output for the period April 2006 to November 2019 with the commodity base denoted as CB as dependent variable and TNI as intensity index for the measurement of ENSO activity. A HAC-Newey-West covariance matrix with Bartlett kernel specification is used. Maximum lags for the dependent variable is 2. Automatic optimal lag structure for the regressors is based on the Akaike Information Criterion (AIC) up to a maximum of 6 lags. Here the independent variables I, W, TNI, IN, GDP, TNI_ELNINO and TNI_LANINA are dynamic regressors and the dummies dum2008, dumELNINO and dumLANINA are fixed regressors. Superscripts *, ** and *** represent the statistical significance at the 10%, 5% and 1% significance level, respectively.*

Independent variables	**Coefficient**	**Std. Error**	**t-Statistic**	**Regression characteristics**	
CB(-2)	0.1040	0.0706	1.4729	R-squared	0.8914
I(-6)	3.8278**	1.5203	2.5178	Adjusted R-squared	0.8698
W(-3)	-6.1651**	2.5547	-2.4132	S.E. of regression	0.2871
IN(-5)	0.1916**	0.0819	2.3408	Sum squared resid.	10.7975
GDP(-1)	0.0871	0.0575	1.5155	Durbin-Watson stat.	2.1144
TNI	0.0168	0.0227	0.7373	Akaike info criterion	0.4964
TNI_ELNINO	-0.0322	0.0777	-0.4149	Schwarz criterion	1.0197
TNI_LANINA	-0.0922**	0.0380	-2.4234	Hannan-Quinn criter.	0.7089
dum2008	0.3412***	0.0996	3.4250		
dumELNINO	0.0693	0.0840	0.8256		
dumLANINA	-0.1223*	0.0651	-1.8784		

Table 19: F-Bounds and t-Bounds test results for the price-effect model - TNI index.

Test results of the F-Bounds and t-Bounds tests on cointegration with respect to the ARDL (2-6-3-6-5-0-4-2) model for the price effect of ENSO on the commodity base measured by TNI as intensity index for ENSO. Per significance level ranging from 10% to 1% the lower and upper bounds as explained in Pesaran, Shin and Smith (2001) provide the critical values against which to test the test-statistics. A value below the I(0) lower bound means there is no cointegration and thus the null hypothesis is accepted. A value above the I(1) upper bound means there is cointegration between the variables and hence the null hypothesis is rejected. A value in between the lower and upper bounds renders the test inconclusive.

F-Bounds Test

	Value	**Signif.**	**I(0)**	**I(1)**
F-statistic	7.3902	10%	1.70	2.83
k	7	5%	1.97	3.18
		2.5%	2.22	3.49
		1%	2.54	3.91

t-Bounds Test

	Value	**Signif.**	**I(0)**	**I(1)**
t-statistic	-4.6485	10%	-1.62	-3.90
		5%	-1.95	-4.23
		2.5%	-2.24	-4.54
		1%	-2.58	-4.88

Table 20: Error Correction Model representation of the price-effect model - TNI index.

*Error Correction Model (ECM) of the ARDL (2-6-3-6-5-0-4-2) regression for the period April 2006 to November 2019 with the commodity base denoted as CB as dependent variable and TNI as intensity index for the measurement of ENSO activity. A HAC-Newey-West covariance matrix with Bartlett kernel specification is used. Maximum lags for the dependent variable is 2. Automatic optimal lag structure for the regressors is based on the Akaike Information Criterion (AIC) up to a maximum of 6 lags. The independent variables I, W, TNI, IN and GDP are the short-term dynamic regressors and the dummies dum2008, dumELNINO and dumLANINA are fixed regressors. The "Δ" in front of each variable stands for the fact that first differences are taken from each variable. The error-correction term (EC) forms the long-term representation of the ECM-equation and its coefficient value (λ) stands for the speed of adjustment to this long-term equilibrium. Superscripts *, ** and *** represent the statistical significance at the 10%, 5% and 1% significance level, respectively.*

Independent variables	**Coefficient**	**Std. Error**	**t-Statistic**	**Regression characteristics**	
Δ (CB(-1))	-0.104	0.0689	-1.5090	R-squared	0.6539
Δ (I)	-8.8155***	0.9247	-9.5338	Adjusted R-squared	0.6063
Δ (I(-1))	-2.5860**	1.1869	-2.1788	S.E. of regression	0.2797
Δ (I(-2))	-0.9304	0.9855	-0.9440	Sum squared resid.	10.7975
Δ (I(-3))	0.0885	0.9910	0.0893	Durbin-Watson stat.	2.1144
Δ (I(-4))	-0.0609	0.9784	-0.0622	Akaike info criterion	0.4078
Δ (I(-5))	-3.8278***	0.9349	-4.0943	Schwarz criterion	0.7954
Δ (W)	4.4991*	2.3718	1.8969	Hannan-Quinn criter.	0.5652
Δ (W(-1))	5.9875**	2.3691	2.5273		
Δ (W(-2))	6.1651**	2.3854	2.5845		
Δ (IN)	0.0040	0.0716	0.0557		
Δ (IN(-1))	-0.2771***	0.0863	-3.2094		
Δ (IN(-2))	-0.2661***	0.0794	-3.3516		
Δ (IN(-3))	-0.1088	0.0773	-1.4068		
Δ (IN(-4))	-0.1916***	0.0729	-2.6284		
Δ (GDP)	0.1470***	0.0326	4.5044		
dum2008	0.3412***	0.1017	3.3549		
dumELNINO	0.0693	0.0459	1.5114		
dumLANINA	-0.1223***	0.0498	-2.4552		
EC(-1)	**-0.2228*** **	**0.0282**	**-7.8918**		

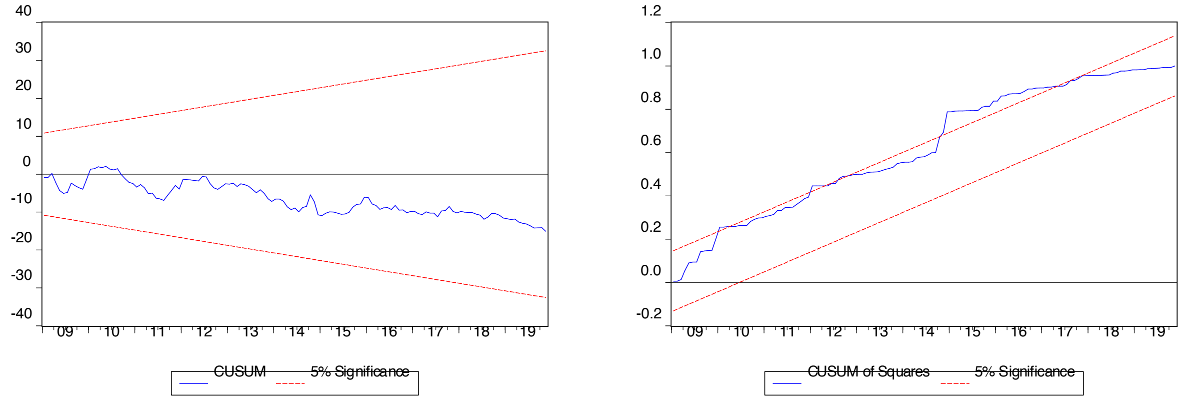
The cointegration equation represented by EC forms the long-run part of the ECM as follows:

$$EC = CB - (-1.7356 * I + 2.3097 * W + 2.1570 * IN + 1.0510 * GDP + 0.0753 * TNI - 0.1447 * TNI\_ELNINO - 0.4138 * TNI\_LANINA)$$

Table 21: CUSUM and CUSUMSQ test results for price-effect model using TNI index.

ARDL-ECM price-effect model stability based on TNI

CUSUM Test | CUSUM of Squares Test



APPENDIX D – Theory of Storage ARDL, Bounds test and ECM output

Table 22: ARDL output of the uncontrolled base model.

*Long-run ARDL (1-6-4) regression output for the uncontrolled base model over the period April 2006 to November 2019 with the commodity base denoted by CB as dependent variable. A HAC-Newey-West covariance matrix with Bartlett kernel specification is used. Maximum lags for the dependent variable is 2. Automatic optimal lag structure for the regressors is based on the Akaike Information Criterion (AIC) up to a maximum of 6 lags. Here the independent variables I and W are the dynamic regressors. Superscripts *, ** and *** represent the statistical significance at the 10%, 5% and 1% significance level, respectively.*

Independent variables	Coefficient	Std. Error	t-Statistic	Regression characteristics	
CB(-1)	0.8124***	0.0654	12.4271	R-squared	0.8598
I(-6)	5.2137***	1.7218	3.0280	Adjusted R-squared	0.8482
W(-4)	5.4659**	2.2754	2.4021	S.E. of regression	0.3100
				Sum squared resid	13.9324
				Durbin-Watson stat	2.0669
				Akaike info criterion	0.5741
				Schwarz criterion	0.8260
				Hannan-Quinn criter.	0.6764

Table 23: F-Bounds and t-Bounds test results for the uncontrolled base model.

Test results of the F-Bounds and t-Bounds tests on cointegration with respect to the ARDL (1-6-4) model for the uncontrolled base model of the theory of storage. Per significance level ranging from 10% to 1% the lower and upper bounds as explained in Pesaran, Shin and Smith (2001) provide the critical values against which to test the test-statistics. A value below the I(0) lower bound means there is no cointegration and thus the null hypothesis is accepted. A value above the I(1) upper bound means there is cointegration between the variables and hence the null hypothesis is rejected. A value in between the lower and upper bounds renders the test inconclusive.

F-Bounds Test

	Value	Signif.	I(0)	I(1)
F-statistic	9.6967	10%	2.17	3.19
k	2	5%	2.72	3.83
		2.5%	3.22	4.50
		1%	3.88	5.30

t-Bounds Test

	Value	Signif.	I(0)	I(1)
t-statistic	-5.3265	10%	-1.62	-2.68
		5%	-1.95	-3.02
		2.5%	-2.24	-3.31
		1%	-2.58	-3.66

Table 24: Error Correction Model representation of the uncontrolled base model.

*Error Correction Model (ECM) of the ARDL (1-6-4) regression output for the uncontrolled base model over the period April 2006 to November 2019 with the commodity base denoted as CB as dependent variable. A HAC-Newey-West covariance matrix with Bartlett kernel specification is used. Maximum lags for the dependent variable is 2. Automatic optimal lag structure for the regressors is based on the Akaike Information Criterion (AIC) up to a maximum of 6 lags. The independent variables I and W are the short-term dynamic regressors. The "Δ" in front of each variable stands for the fact that first differences are taken from each variable. The error-correction term (EC) forms the long-term representation of the ECM-equation and its coefficient value (λ) stands for the speed of adjustment to this long-term equilibrium. Superscripts *, ** and *** represent the statistical significance at the 10%, 5% and 1% significance level, respectively.*					
Independent variables	**Coefficient**	**Std. Error**	**t-Statistic**	**Regression characteristics**	
Δ (I)	-7.9154***	0.9435	-8.3893	R-squared	0.5535
Δ (I(-1))	-1.7255*	1.0038	-1.7189	Adjusted R-squared	0.5231
Δ (I(-2))	-2.0395**	1.0042	-2.0309	S.E. of regression	0.3079
Δ (I(-3))	-1.2783	1.0477	-1.2201	Sum squared resid.	13.9324
Δ (I(-4))	-1.1821	1.0626	-1.1125	Durbin-Watson stat.	2.0669
Δ (I(-5))	-5.2137***	0.9766	-5.3386	Akaike info criterion	0.5487
Δ (W)	5.8324**	2.6025	2.2411	Schwarz criterion	0.7620
Δ (W(-1))	4.3636*	2.4951	1.7488	Hannan-Quinn criter.	0.6353
Δ (W(-2))	4.7072*	2.4999	1.8829		
Δ (W(-3))	-5.4659**	2.6157	-2.0897		
EC(-1)*	**-0.1876*****	**0.0345**	**-5.4306**		

The cointegration equation represented by EC forms the long-run part of the ECM as follows:

$$EC = CB - (2.1266 * I + 6.1193 * W)$$

Table 25: CUSUM and CUSUMSQ test results for the uncontrolled base model.

ARDL-ECM theory of storage uncontrolled base model stability (CB)

CUSUM Test

CUSUM of Squares Test