Norwegian University
of Life Sciences

Master's Thesis 2023 30 ECTS
School of Economics and Business

Can the Oil Price Predict the Wheat Price?

Atle Torgersen
Applied Economics and Sustainability

Can the oil price predict the wheat price?

Abstract

This paper examines if the oil price can help predict the wheat price. The analysis is based on the commodity prices of wheat and oil, traded in the United States but which is generalizable to represent their respective global markets. The study consists of monthly observations, spanning from January 1990 to April 2023, divided into five different time intervals. I compare the outcomes of a textbook version of the univariate autoregression model with a nonlinear specification model that extends Hamilton's framework as proposed by Hamilton (2003) and subsequently expanded upon by Hamilton (2009), which bears resemblance to an ARDL form. I employ the BIC test to determine the optimal lag numbers for both models and utilize granger-causality test to ascertain whether the oil price can, in fact, help predict the wheat price. My main hypothesis is that the time series of oil does not granger-causes the time series of wheat, my findings are in line with the literature and the finds from Umar Z. , Gubareva, Naeem, & Akhter (2021), where I reject the null hypothesis to conclude the oil price, in fact, does help predict the wheat price at the 5% significance level. However, this outcome does not accurately reflect the time periods extending beyond the start of 2021, in which I acknowledge the null hypothesis.

The data covers a span of three decades, encompassing significant events, where the worldwide covid-19 pandemic and Russia's invasion of Ukraine, which hold special relevance to my findings. The United Nations recognizes the significance of food and energy by incorporating them in the UN's sustainability goals of achieving "zero hunger" (United Nations Sustainable Development Goals, 2023) and ensuring "clean and affordable energy for everybody" (United Nations Sustainable Development Goals, 2023). However, it is worth noting that these objectives may potentially conflict with each other. Policy efforts like the Renewable Fuel Standard in the United States of 2005 demonstrate the political will to cut emissions (United States Environmental Protection Agency, 2023).

Acknowledgement

The process of writing my master thesis has taught me a great deal about myself, it has been humbling, frustrating, overwhelming, challenging, interesting, fun, rewarding and educational. Despite all the time and hard work I have put into writing my thesis, I consider myself privileged for having the opportunity to pursue a master's degree in economics and sustainability, and I am truly grateful for all the knowledge I have acquired. I have put theory into practice, enhanced my analytical skills and further developed my understanding for coding in Python, as well as maturing my independent and critical thinking in general.

I would like to extend a sincere thank you to my supervisor, Kine Josefine Aurland-Bredesen, for her prompt and honest feedback as well as her invaluable insight, guidance, and encouragement throughout this process.

Lastly, I would like to thank my family, friends, fellow students, professors and faculty for their unwavering support, positivity and understanding throughout my master's degree and the process of writing my thesis in particular.

Contents

List of Figures Chapter two

List of Figures Chapter four

List of tables

1. Introduction

This thesis is comprised of six main chapters. The initial chapter, known as the Introduction, serves the purpose of providing motivation for my research question, theory, and methodology. This is achieved by drawing upon existing literature and establishing the position of this thesis within the broader context of the existing literature. The second chapter is Background which is divided into five subchapters. This chapter aim to provide a comprehensive overview of the oil and wheat market, significant major events which fall inside the research period as well as further exploring the United Nations sustainability goals and the renewable fuels standard which was implemented in the United States in 2005. The third chapter of the thesis focuses on Economic Theory, Mechanisms, and Literature Review. The first subchapter provides a comprehensive review of relevant literature. The second subchapter explores two mechanisms through which the oil price may impact the wheat price, namely oil as an input factor in wheat production and biofuel as a substitute for petroleum-based fuels. The third subchapter presents the economic theory of supply and demand, utilizing figures and examples to support the mechanisms discussed above. The fourth chapter, Method and Data, is divided into two primary subchapters. The initial subchapter focusses on presenting the data utilized in this thesis, whereas the subsequent subchapter will offer a comprehensive explanation of the Method employed. In the fifth chapter, I present my research findings, which encompass the outcomes derived from my models. The results for each model are presented for every time interval and lag length. Furthermore, a comprehensive collection of graphs is provided to help illustrate my findings. The sixth chapter, Results, is devoted to the discussion, to explaining the limitations of this thesis, making suggestions for further research as well as providing a summary.

The primary motivation behind my decision to explore if oil prices can help predict the future wheat prices stems from my desire to offer empirical evidence that can assist aid organizations, governments, farmers, and investors in formulating necessary strategies to adapt to evolving circumstances. Given that approximately 20% of the total caloric and protein intake of the global human population is derived from wheat, it becomes evident why this agricultural commodity holds significant importance (Shiferaw, et al., 2013). Recent events as the supply constraints due to the global covid-19 pandemic, shortly followed by Russia's invasion of Ukraine, where war efforts impeded exports of grain from Ukraine and sanctions impeded export of oil from Russia, causing turbulence in the global oil and grain markets.

There is an extensive body of literature that explores the interconnectedness between the energy and agricultural sectors. To the best of my knowledge, no paper has utilized the framework developed by Hamilton, which I describe in chapter four, there is however a resemblance between his nonlinear specification model and the ARLD model utilized by Rafiq & Bloch (2016). Rafiq & Bloch (2016) investigates links between the oil price and agricultural commodities and metals, they conclude that oil prices do not have a large impact on cereal prices as a cluster, however they find that a decline in oil prices is significant at the 5% level for a reduction in wheat prices. I have made significant contributions to the existing body of literature through two distinct avenues. Firstly, I have employed Hamilton's framework, which sets my research apart from previous studies. Secondly, I have utilized a dataset that spans monthly observations up until April of 2023. This contrasts with Rafiq & Bloch (2016), who rely on annual data only up to 2011, which due to recent events might be considered somewhat outdated. The disparity in our findings may be attributed to the contrasting time periods and frequency of data used in our respective analysis.

While the relationship between the oil price and the wheat price remains a topic of debate in the literature, many studies indicate the existence of granger-causality at the 5% significance level in at least one specific case or scenario mentioned in their respective papers. Recent research by Umar et. al. (2021) concludes that oil shocks granger-causes wheat, grains, and live cattle at the 5% significance level.

My aspiration is that my thesis will contribute to the existing body of literature and hold significance for policymakers, aid organizations, farmers, and other relevant stakeholders. My objective is to bring clarity to the potential conflict between two sustainability goals set by the United Nations, namely "zero hunger" (United Nations Sustainable Development Goals, 2023) and "clean and affordable energy" (United Nations Sustainable Development Goals, 2023). These goals may clash due to political implementations, such as the Renewable Fuel Standard in the United States, which requires a specific blend of biofuels in petroleum-based fuels (United States Environmental Protection Agency, 2023). The clash can arise when agricultural crops, which may be used for food, instead is used for fuel, this can help reduce emissions, but potentially limit the food supply, as crop land can be considered to be a scarce resource.

2. Background

This chapter aims to provide a comprehensive overview of the oil and wheat market, based on existing literature and trustworthy sources, to help substantiate my research question, theory, and method. The chapter is divided into four subchapters, each focusing on different aspects. The first subchapter explores two UN sustainability goals that may potentially conflict with each other. The second subchapter discusses the renewable fuels standard, which was initially implemented in the United States in 2005. The third subchapter summarizes significant global events that could have an impact on the oil and wheat markets. The fourth subchapters delve into the specifics of the wheat market.

2.1. Potentially conflicting UN Sustainability goals

The second UN goal is "Zero Hunger", the number of people experiencing hunger has increased in the years after 2015, according to (United Nations Sustainable Development Goals, 2023). While the seventh UN goal is "Affordable and Clean Energy" where they point out that although electricity generation is getting cleaner (around 30% of electricity is produced from renewable sources, the same trend is not apparent for transportation and cooking, which implies there is still a long way to go to transition from fossil fuels to renewable energy worldwide (United Nations Sustainable Development Goals, 2023).

The objective of UNs sustainability goal of zero hunger may be addressed through affordable and accessible food. However, this objective might clash with the goal of affordable and clean energy, especially if biofuels are considered as part of the clean energy solution. This potential conflict arises if we discover that the price of oil has a direct impact on the price of wheat. This topic is occasionally referred to as the "food-fuel" debate in the existing literature Umar et. al., (2021).

2.2. The Renewable Fuel Standard

The Renewable Fuel Standard, RFS, was first created in 2005 and further expanded in 2007, it is administered by the United States Environmental Protection Agency, EPA, and collaborates with the United States Department of Agriculture and the United States Department of Energy. The Renewable Fuel Standard was created to ensure a minimum of renewable fuels to either reduce or replace traditional petroleum fuels for transportation and heating (United States Environmental Protection Agency, 2023).

Chen, Kuo, & Chen (2010) analyses the interconnectedness, or spillover effects, from biofuels based on ethanol, mainly produced from soybeans and corn. Increased demand for soybeans and corn impacts the wheat price in two main ways, the first is due to the change of the limited crop land use from food to biofuels production, the second is through the substitution effect where wheat, in part, is a substitute for soybeans and corn. This increases the demand for wheat, while reducing, or at least restricting, the supply of wheat, which in turn could increase the wheat prices, Chen et. al. (2010).

2.3. Major world events which have the potential to affect the oil or wheat market

This subchapter provides an overview of major events which could potentially influence the global oil and wheat markets, which are the commodities of interest in this thesis. The knowledge of, and from these events, has been utilized in the construction, analysis, and result section to validate the outcomes and provide a rationale for the results.

2.3.1. Economic Recessions in the United States

Economic Recessions in the United States are of interest for two main reasons. Firstly, since both the wheat and oil commodities I use in this thesis are traded in the United States, which I come back to in chapter fours. Secondly, given its status as the largest economy in the world, the American economy holds significant importance for the global economy.

There are different ways to determine if an economy is in a recession, as well as when the recession begins and ends, respectively. However, in the United States the Business Cycle Dating Committee of the National Bureau of Economic Research is considered authoritative in deciding when the United States is in a recession, as well as the length of said recession. The committee leverage several pivotal measures to analyze the aggregated real economic activity, as the nonfarm payroll (jobs outside the agricultural sector), industrial production, real personal consumption expenditure (PCE), real personal income less transfers (PILT), employment as measured by the household survey and the wholesale-retail sales adjusted for price changes, to determine the date of the peak and the trough. The peak is a month where the economic indicators listed above hit their highest levels before they start to decline, as opposed to a trough, which is when the economic indicators hit their lowest point, before they start to increase. Although the decision to define a month as a peak or trough is based on economic indicators, the committee does enjoy some discretion in their decision-making process (National Bureau of Economic Research, 2022).

According to the US Business Cycle Expansion and Contractions chronology, there has been four recessions in the United States from 1990 through to its update as of March 2023. The recessions are listed from their start to end date as follows: July 1990 to March 1991, March 2001 to November 2001, December 2007 to June 2009 and February 2020 to April 2020 (National Bureau of Economic Research, 2023).

2.3.2. Major world events

In addition to economic recessions as discussed above, other major events since 1990 which may impact the oil and or wheat markets, in ascending order are, the collapse of the Soviet Union, the oil price collapse in 2014 (Stocker, Baffes, & Vorisek, 2018), the global Covid-19 pandemic in 2020, and Russia's invasion of Ukraine in 2022, all have the potential to temporarily or permanently change the global oil or wheat markets.

2.4. The wheat market

The objective of this subchapter is to point out pertinent details regarding the wheat market, while also presenting a comprehensive overview of global wheat exports.

2.4.1. Relevant background information on the wheat market.

According to Shiferaw, et al., (2013) wheat is essential for global food security as it accounts for roughly 20% of calories and proteins consumed by the human population worldwide. They also state that around 20% of the wheat production is used to feed livestock, further cementing wheats important role as a food and as an input factor for meat production (Shiferaw, et al., 2013).

Enghiad, Danielle Ufer, & Thilmany, (2017) investigate the established link in excising literature between oil-, stock- and global wheat prices to investigate how this affects food security. They argue wheat is important in humans' diet and therefore, especially in the context of climate change, is important to investigate further. They use these key variables (oil, stock, and wheat commodity) from the five largest wheat exporting countries (or regions) to further investigate this connection. Their study period spans from 1980 to 2013. They find that, although aid distorts the market somewhat, there are different wheat prices in different countries or regions.

According to the U.S. Standards published by the United States Department of Agriculture, wheat is divided into eight classes, Hard Red Winter wheat, Hard Red Spring wheat, Soft Red Winter wheat, Durum wheat, Hard White wheat, Soft White wheat, Unclassed wheat, and Mixed wheat. While

Durum wheat, as an example, is divided into three subclasses, Hard Red Winter wheat has no subclasses (United States Department of Agriculture, 2014). Please note that the wheat price I use in this thesis is based on Hard Red Winter wheat, traded in the United States (CME Group, 2019).

2.4.2. Global wheat exports

Due to limited publicly accessible graphs pertaining the global wheat market, I have gathered data on global wheat export from the United States Department of Agriculture – Foreign Agricultural Service (United States Department of Agriculture - Foreign Agricultural Service, 2023) to create the graphs presented in this subchapter. Moving forward I will refer to data from this source as USDA which is the official abbreviation of the United States Department of Agriculture. Further information on the collection and handling of this data is described in appendix 1 attached below. Please note the top exporting countries over this time interval, in descending order, is the United States of America, the European Union, Canada, Australia, Russia, Argentina and Ukraine, which in total accounted for about 85% of the volume of global wheat exports, which is visualized in the graph below.

I take a closer look at global wheat exports from 1990 to 2022. We use data from USDA to shed light on the largest exporting countries, as well as global export trends.

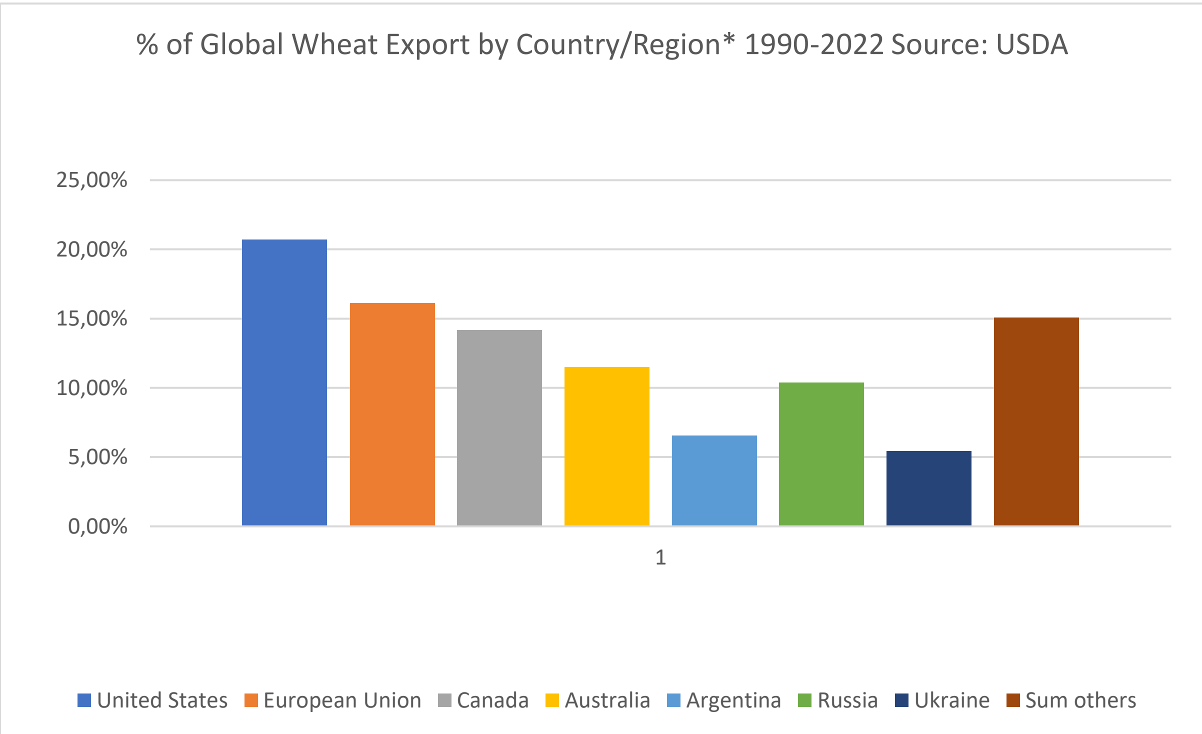


Figure 2. 1. Global wheat export in %

From figure 2.1., above, we notice that the seven largest wheat exporting countries, in descending order, are the United States of America (roughly 21%), the European Union (roughly 16%), Canada (roughly 14%), Australia (roughly 12%), Russia (roughly 10%), Argentina (roughly 7%) and Ukraine (roughly 5%), which adds up to roughly 85% of all global wheat exports. All other wheat exporting countries is accounted for in the "sum other" at about 15% from 1990 to 2022 (United States Department of Agriculture - Foreign Agricultural Service, 2023).

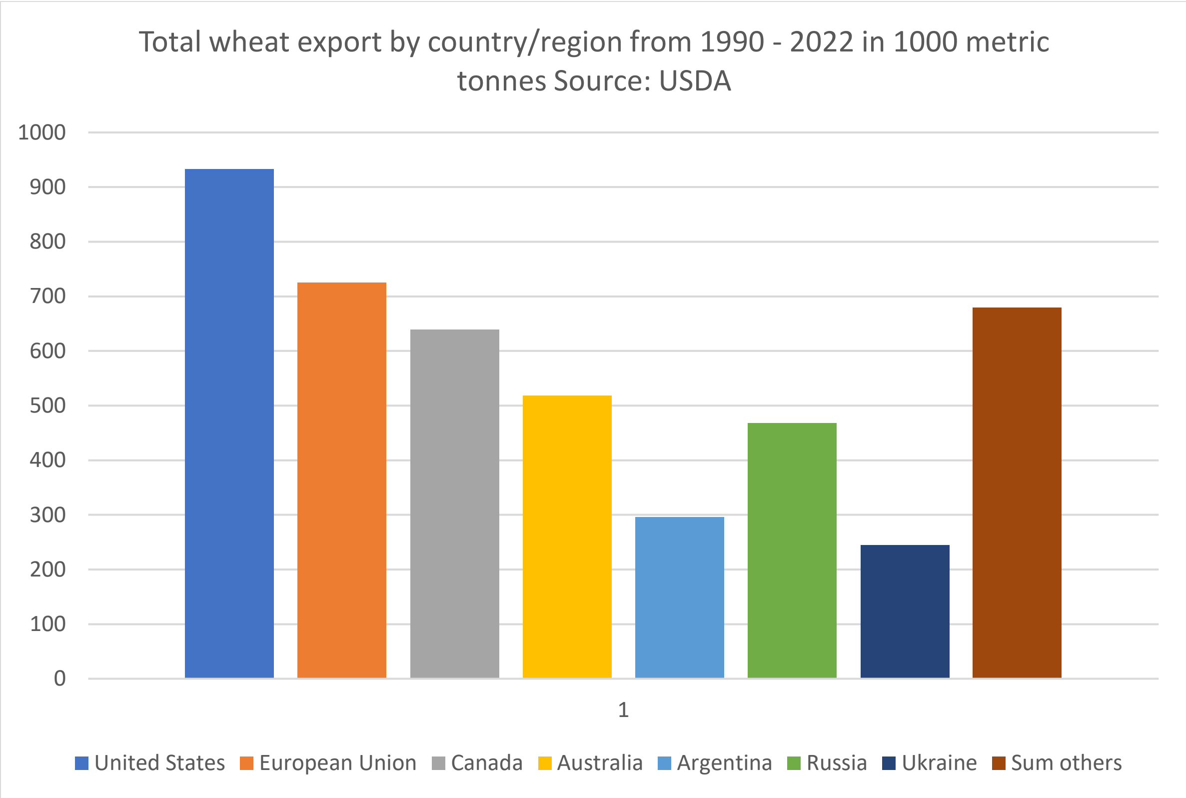


Figure 2. 2. Total wheat export in 1000 metric tonnes

Figure 2.2., is presented in 1000 tons of wheat exports from 1990 to 2022. The seven largest exporters, in descending order are the United States (933,173,000 tons), the European Union (725,739,000 tons), Canada (639,180,000 tons), Australia (518,655,000 tons), Russia (468,418,000 tons), Argentina (295,969,000 tons), Ukraine (244,439,000 tons), and all other wheat exporting countries is accounted for in the "sum other" (680,087,000 tons). These countries exported 3,825,537,000 tons in total from 1990 to 2022 (United States Department of Agriculture - Foreign Agricultural Service, 2023).

I would like to draw your attention to figure 2.3., for an overview of the development of global wheat exports, by country, from 1990 to 2022. The United States is the largest exporter in the first half of the period, while Russia becomes the largest exporter towards the end of the period.

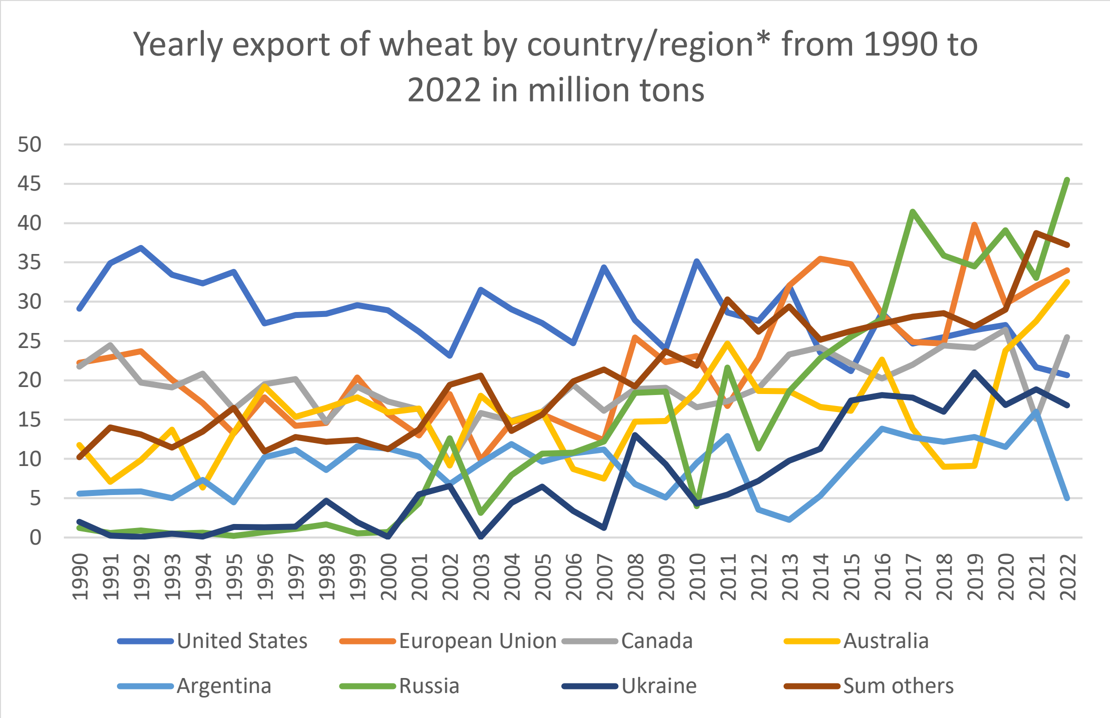


Figure 2. 3. Development of wheat export by country from 1990 to 2022

Figure 2.4. below, visualizes the total volume of global wheat export by year from 1990 to 2022. The yearly export has an upward trend, and the volume doubles from around 100 million tons in 1990 to above 200 million ton in 2022.

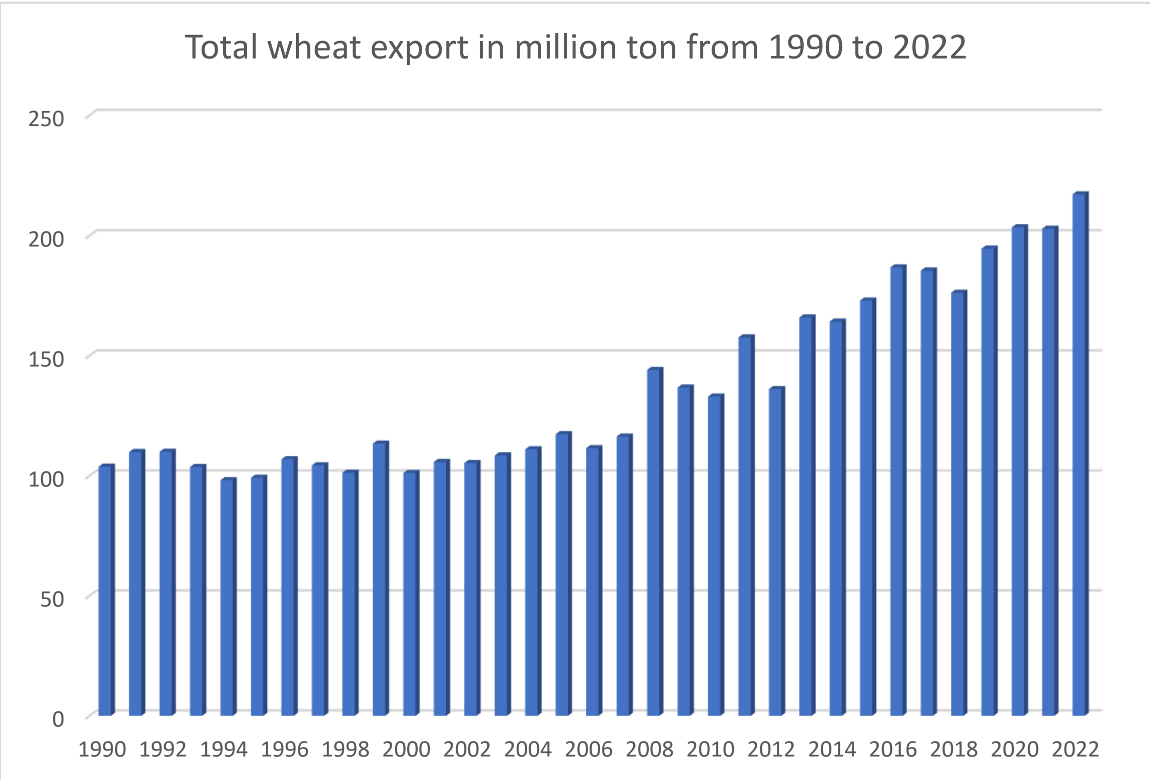


Figure 2. 4. Development of global wheat export from 1990 to 2022

3. Economic Theory, Mechanisms and Related Literature

This chapter is divided into three main subchapters. The first subchapter contains a review of related literature on the link between oil and the wheat price, as well as establishing where this thesis fits into the existing research in this field. The second subchapter delves into two mechanisms in which the oil price could affect the wheat price, the first mechanism is oil as an input factor in wheat production, whereas the second mechanism is biofuel as a substitute for petroleum-based fuel. The third subchapter presents the economic theory of supply and demand, illustrated with figures and examples to substantiate the mechanisms described in the second subchapter.

3.1. Related Literature

There is no clear consensus on whether the oil price has a causal link to the wheat price in the existing literature, however, many papers prove granger-causality at the 5% significance level in some or all instances they investigate. Explanations of the inconclusiveness could be due to, for example, differences in the method, data, frequency, time interval, or a combination, as to why different papers reach different conclusions.

Umar et. al., (2021) use a multi-variable VAR (vector autoregressive) method to analyze the interconnection of oil price shocks, agricultural commodity returns as well as volatility by utilizing their approach of shock construction. Their study period is from January 2002 through July 2020, they use monthly data from S&P GSCI Agricultural Indexes, Crude Oil Futures from the New York Mercantile Exchange, the VIX index, and the World Integrated Oil and Gas Producer Index from Datastream. They use 14 variables of which 11 are agricultural commodities, such as: cocoa, cotton, coffee, feeder cattle, live cattle, lean hogs, livestock, grains, sugar, soybeans and wheat, the crude oil futures with one month to maturity as well as the VIX index to measure risk perception and the World Integrated Oil and Gas Producer Index from Datastream to proxy oil producing firms. Firstly, they find that oil shocks have a statistically significant Granger-Causality on grains, live cattle, and wheat. Secondly, they find that livestock is the largest transmitter of spillovers of price and volatility while lean hogs is the largest receiver Umar et. al., (2021).

Lu, Yang, & Liu, (2019) investigate potential spillover effects from the crude oil market to the agricultural commodity market by a bivariate heterogeneous autoregressive model. Their study uses daily prices, from the days in which crude oil, corn, soybean, and wheat are traded, spanning from the 1st of July 2008 to the 29th of December 2017. The authors conclude there is a bidirectional volatility spillover from the oil to the wheat market, at the 5% significance level when using the Granger-Causality test in the crisis period (2008-2009) but not in the post-crisis period. They attribute their findings to the financialization of the agricultural market leading up to the crisis period, as opposed to agricultural subsidies in the post-crisis period.

Rafiq & Bloch (2016) investigate the linkages between oil and commodity prices as cereals or metals, by using nonlinear and linear ARDL models to identify short- and long run relationships. Their study uses annual data from 1900 to 2011 of 26 commodities. The authors conclude oil prices do not have a large impact on cereal prices as a cluster, however they find that a decline in oil prices is significant at the 5% level for a reduction in wheat prices.

Mensi, Hammoudeh, Nguyen, & Yoon (2014) use two multivariate GARCH models, the BEKK-GARCH and DCC-GARCH model, firstly in their attempt to quantify spillover effects from four oil markets to four cereal markets, and secondly examining if announcements by OPEC offer significant impacts on spillover effects. Their study uses spot prices (daily closing) from the 3rd of January 2000 to the 29th of January 2013, on their eight variables, of which four relates to the oil markets (WTI oil, Europe Brent oil, gasoline, and heating oil) and four related to the cereal markets (barley, corn, sorghum, and wheat). They conclude with a unidirectional causality from WTI to wheat (as well as corn and barley).

Reboredo (2012) use copulas (Archimedean) to investigate co-movements of the price of oil, soybean, wheat, and corn, leveraging weekly data spanning from January 1998 to April 2011. He concludes there is no contagion from the crude oil to the agricultural markets, indicating neutrality on the agricultural markets based on price changes for crude oil.

Nazlioglu (2011) use the approach of Toda-Yamamoto and method of Diks-Panchenko to test for linear and nonlinear granger-causality, respectively. He uses weekly price data from 1994 to 2010 for oil, corn, soybean, and wheat. The author concludes that there is no granger-causality between the oil and agricultural commodities in the linear model, whereas there is nonlinear causal links for the nonlinear model, in addition there is strict causality for the nonlinear model from oil to corn and soybean prices.

Despite a substantial body of academic research on the interconnectedness of energy commodities (as oil) and food commodities (as cereals), to the best of my knowledge, no other study has applied Hamilton's framework to analyze said interconnectedness between the oil- and wheat price. Please note that said framework is based on Hamilton (2003) and builds on Hamilton (2009), which I describe in the next chapter. Furthermore, to the best of my knowledge, there are no other studies with as current data as I used in the analysis of this thesis, current data is especially relevant due to current events as the global pandemic and Russia's invasion of Ukraine since 2020. My thesis addresses these voids in the existing literature.

3.2. Mechanisms

This subchapter delves into relevant literature to substantiate the two mechanisms in which the oil price could affect the wheat price. The first mechanism considers oil as an input factor in wheat production, where an increase in the oil price in turn increases the cost of producing wheat. The second mechanism looks at biofuel as a substitution for petroleum-based fuel, where the substitution is determined by the farmer, who must decide to produce agricultural crops for food or fuel purposes.

3.2.1. Oil as an input factor in wheat production

Umar et. al. (2021) also discusses three mechanisms which could explain the co-movement of agricultural prices and energy prices. The first mechanism is the "food-crisis" which they argue was a result of higher crude oil prices, the second mechanism is the substitution effect from agricultural food production to agricultural biofuel production, whereas the third mechanism is oil dependent input factors in agricultural production, for example for fertilizer, transport and for machinery in agricultural production.

Piringer & Steinberg (2008) use life-cycle assessments to analyze the energy budget of wheat production in the United States. Their data is representative for wheat production in the seventies and eighties in the U.S., and they conclude that diesel fuel in wheat production accounts for 25% of the total energy input of wheat production.

Wheat producers, like any agricultural producer, incur a range of different production costs. According to USDA ERS, wheat producers spend approximately 8-10% of operating costs, or 3-4% of total production costs (where overhead costs are included), for fuel in wheat production. (United States Department of Agriculture - Economic Research Service, 2023).

Furthermore, the cost of fertilizers account to approximately 35% of operating costs, or approximately 14% of total production costs (United States Department of Agriculture - Economic Research Service, 2023). According to Chen, Chang, Chen, & McAleer (2012) several of the global fertilizer prices are in fact, influenced by the oil price.

3.2.2. Biofuel as a substitute for petroleum-based fuel

Baumeister & Kilian (2014) use bivariate autoregressions to investigate if there is a pass-through from oil prices to food prices because of the biofuel policy implemented in the U.S. in 2006. They find that there is no pass-through of prices from oil to food prices, however, they conclude there is a significant pass-through effect from oil to agricultural commodity prices as for example wheat, soybean, and corn.

Saghaian (2010) argue that the oil- and agricultural markets are becoming more interconnected due to the oil-ethanol-corn link. This is a result of more biofuels being due to government mandates as biofuels are mixed with refined oil and sold as petrol (diesel/petrol/benzin). Wheat is primarily affected as farmland is a limited resource, and increased demand for biofuels makes some farmers substitute wheat productions for soybeans or corn which is used to produce biofuels.

Chen, Kuo, & Chen (2010) use weekly data on oil prices, and grain prices (wheat, corn, and soybean) from the beginning of 2005 to mid-2008. Their look into the soybean-based bio-diesel and corn-based ethanol production, and their goal is to investigate any links between the global oil and grain prices, they find that changes in grain prices are significantly influenced by the changes in oil prices.

Paris (2018) use a nonlinear regression model based on cointegrating smooth transition to investigate the long-term effects of oil prices on agricultural commodity prices, he uses corn and soybean in the U.S. and rapeseed in Europe, as well as corn and sunflower to investigate if biofuels affect the substitute effect from food to fuel farming. He uses monthly data and concludes that increased biofuel production contributes to price increases in agricultural commodities.

3.3. Economic Theory

This subchapter on economic theory is divided into two main parts. The first part offers relevant background information on shifts and movements along the demand or supply curve. The second part focuses on substantiating the interconnectedness of the oil and wheat markets through the substitution mechanism and the input factor mechanism, through examples with figures.

3.3.1. Shifts and movements along the demand and supply curve

A market for goods or services is in equilibrium when the supply matches the demand. The state of equilibrium can be temporarily challenged by movements along the supply or demand curve due to price level variations, or permanently challenged due to shifts in the supply or demand curve caused by changes in the market fundamentals.

There are several key factors that influence demand, including the price of the good or service, the price of complementary goods, income level (for the consumers), tastes and preferences, future price expectations, the number of consumers, and the distribution of income (Hansen, 2013).

Similarly, the supply of a good or a service is influenced by factors such as its price, the cost of input factors, technological advancements, government-imposed fees or subsidies, future price expectations, as well as the number of suppliers (Hansen, 2013).

3.3.2. The input factor mechanism – technological innovation in the oil sector

The first mechanism helps establish the link between the oil- and wheat price, based on oil as an input factor for wheat production. Figure 3.1., below illustrates an example of a technological innovation in the oil sector (e.g., fracking), which shifts the supply curve of oil outwards. The implications of an outwards shift of the supply curve are a reduction in the equilibrium price and an increase in the equilibrium quantity, notated as P_1^* and Q_1^* and for the price and quantity before the shift in supply and P_2^* and Q_2^* after the shift in supply, respectively. For this analysis, I assume the oil price reduction led to reduced prices for petroleum-based fuels in my analysis.

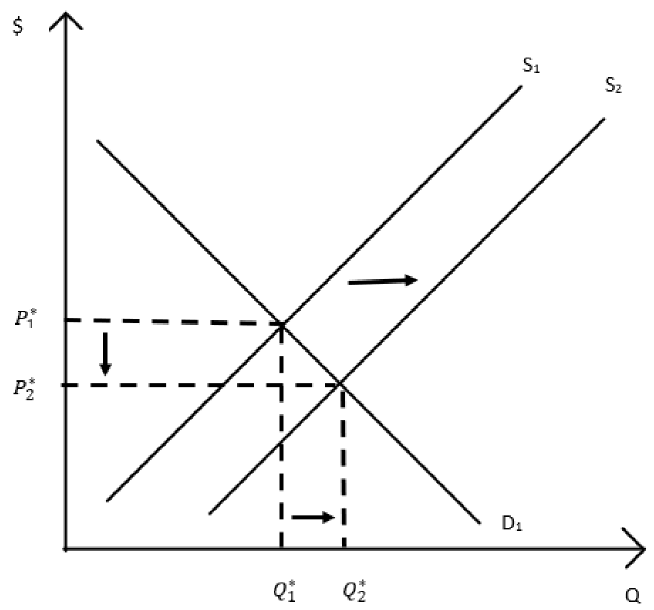


Figure: 3.1. Price and quantity changes due to an outward shift in the supply curve, due to a technological innovation in the oil market.

The technological innovation in the oil market reduces the cost of petroleum-based fuel, which in turn reduces the price of an input factor for wheat producers. Remember that the wheat market is characterized as a free market, and therefore we assume all cost savings will benefit the consumers. Reduced input costs are likely to shift the supply curve of wheat outwards, illustrated in figure 3.2., below. As input prices are reduced, the supply curve shifts outwards, which in turn lead to a lower equilibrium price and a higher equilibrium quantity, notated as P_1^* and Q_1^* for the price and quantity before the shift in supply, and P_2^* and Q_2^* after the shift in supply, respectively.

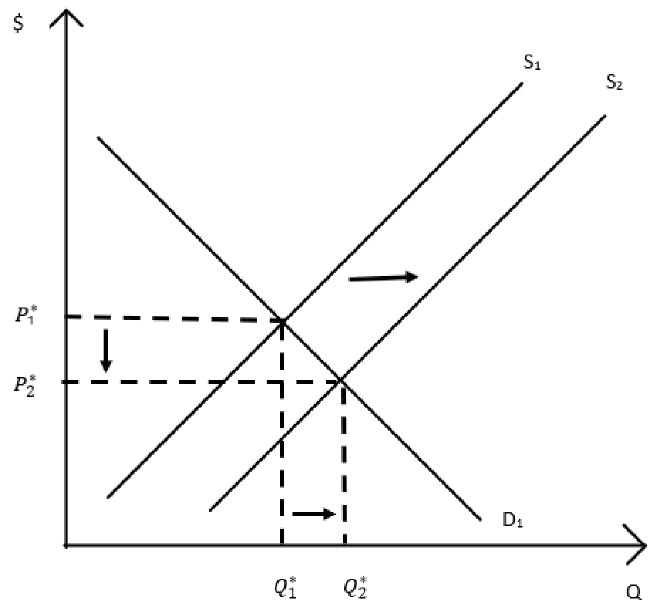


Figure: 3.2. price and quantity changes due to an outward shift in the supply curve, due to reduced input costs.

3.3.3. The substitution mechanism

The second mechanism looks at the substitution effect within the agricultural sector, where farmers face the choice of producing crops for food or biofuel production. If a farmer can make more money from growing agricultural crops for biofuel production than food production, it is safe to assume the farmer will substitute away crops for food production in favor of crops for biofuel production. Whereas wheat is mainly used for food production, soybeans and corn could be used for either food or biofuel production. In this context, I assume that available land for agricultural production is scarce, and the substitution mechanism is based on farmers choosing to produce soybeans and corn for biofuel production instead of wheat for food production. Please note that biofuel can help reduce emissions, either as a stand-alone alternative to fossil fuels or as a blend into fossil fuel to cut back on emissions.

The primary effects of the new government policy

A government policy requires biofuel to be blended with fossil fuel to cut emissions. I assume that available biowaste is insufficient to meet the new increased demand, and the difference must be made up from farmers growing crops specifically for biofuel production. The new policy will likely shift the demand curve for biofuels outwards (soybean and corn), figure 3.3., below help illustrate the likely shift in the demand curve for soybean and corn. An outward shift in the demand curve will likely lead to an increased equilibrium price and equilibrium quantity, notated as P_1^* and Q_1^* for the price and quantity before the shift in demand and P_2^* and Q_2^* after the shift in demand, respectively. Please note, an increase in biofuel blend to fossil fuel could reduce demand for fossil fuels, which in turn could reduce the demand for fossil fuels, lowering the price of fossil fuel.

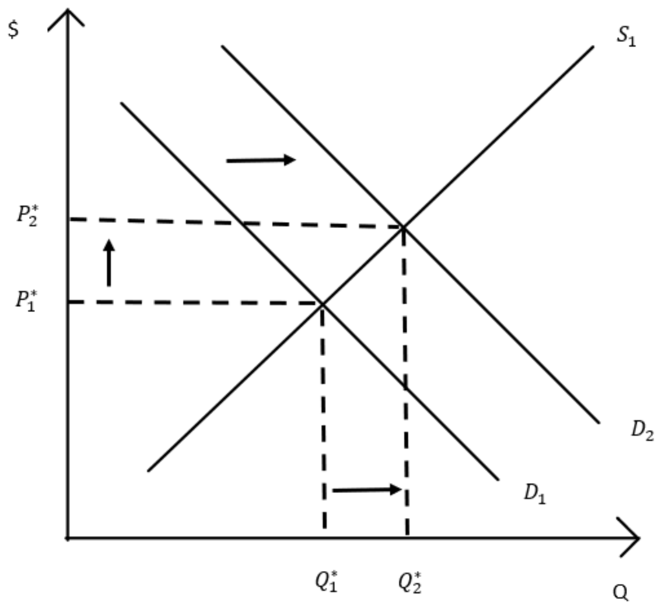


Figure: 3.3. Price and quantity changes due to an outward shift in the demand curve for soybean and corn due to government policy.

The secondary effects of the new government policy

As more arable land is prioritized for biofuel production (soybean and corn), there is less land available for producing other crops, such as wheat. A possible secondary implication of the government policy is an inward shift of the supply curve for wheat. Figure 3.4., below helps illustrate the likely impacts from the policy in the wheat market. An inward shift in supply lead to an increased equilibrium price and reduced equilibrium quantity, notated as P_1^* and Q_1^* for the price and quantity before the shift in supply and P_2^* and Q_2^* after the shift in supply, respectively.

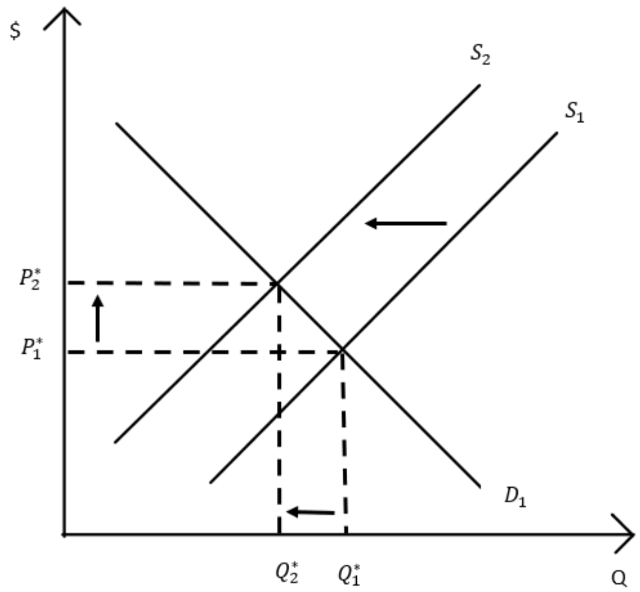


Figure: 3.4. Price and quantity changes due to an inward shift of the supply curve for wheat as a secondary effect of government policy.

The two mechanisms described in this subchapter are analyzed using fundamental economic theory on supply and demand. The example used to illustrate the substitution mechanism clearly underscores the importance of government analyzing new policy before implementing to avoid adverse secondary effects.

4. Method and data

This chapter is divided into two subchapters. The first subchapter is about the data I have used, including graphs and statistics of said data. The second subchapter is about the method I have used to analyze the data as well as a description of the statistical tests I have carried out to ensure my results are statistically valid at the 5% statistical significance level.

Throughout this thesis, but especially when dealing with the data, method, and results (the results will be presented in the next chapter) there are made certain assumptions. We assume that normal economic principles apply; as free global trade, where supply versus demand equilibrium determines the price, that all actors have perfect information (especially regarding the type and quality of any commodity), that market participants behave rationally to maximize profits and minimize costs and that there are no transaction costs.

4.1. Data

The data I have used in my thesis is downloaded from The Federal Reserve Fund of St. Louis webpages. This is due to the Federal Reserve Funds easy to access and easy to use data for time series as they offer downloads in CSV format with American formatting (points as decimal separator and comma as thousand separator) which makes these time series suitable for analysis in Python. The author of the oil price index is the United States Bureau of Labor Statistics (U.S. Bureau of Labor Statistics, 2023) and the author of the wheat price is the International Monetary Fund (International Monetary Fund, 2023). Please note that the International Monetary Fund, in turn, refers to the United States Department of Agriculture (USDA) as their source for the Hard Red Winter wheat prices in their report on Technical Documentation under Primary Commodity Prices on their website (International Monetary Fund, 2019)

The observations are reported monthly for each time series and the datasets start in January 1990 and end in April 2023 (1990-01-01:2023-04-01), which gives us a total of 400 observations for each time series. The oil price is originally an index which is set to equal 100 in 1982 (the average of the year), I have reindexed this time series to equal 100 in January 1990. The global wheat price is reported in USD and this time series has been indexed to equal 100 in January 1990. None of the time series are seasonally adjusted or adjusted for inflation.

4.1.1. Oil Prices

As mentioned in subchapter *4.1. Data*, the data for the Oil Price Index is retrieved from the Federal Reserve Bank of St. Louis, henceforth abbreviated to FRED, which specify their source for the time series in question to be the United States Bureau of Labor Statistics. The oil price index I use is named WPU0561 and is the producer price index (PPI) of crude petroleum in the United States of America. It consists of monthly observations from January 1947, and is still published every month, as the most recent observation was published on the 13th of July 2023 for the month of June 2023 (U.S. Bureau of Labor Statistics, 2023). The unit of this time series is an index which is set to equal 100 in 1982, please note that the average of the observations for 1982 equals 100. I have reindexed the oil price index to equal 100 in January 1990 to be in line with the time series of global wheat prices in my analysis. The data is not seasonally adjusted and consists of a total of 400 monthly observations from 1990-01-01:2023-04-01 (U.S. Bureau of Labor Statistics, 2023).

We have reason to believe that the oil price described above, which is a producer price index for the United States of America, provides us an accurate understanding of the crude petroleum prices the wheat industry face, both directly through the cost of energy for production and indirectly through transportation and fertilizer costs. Please note that other potential time series for oil prices as WTI- or Brent Crude are typically more volatile than producer price indexes. During Covid-19 the WTI Crude oil price was observed with negative prices at certain points (as the 20th of April 2020) due to a shock in global oil demand as well as full storage capacity Kubursi (2021).

Figure 4. 1 Oil Price Index

The graph above shows the producer price index of the oil price from January 1990 to April 2023 in nominal terms (remember the time series is indexed), based on the data from the U.S. Bureau of Labor Statistics.

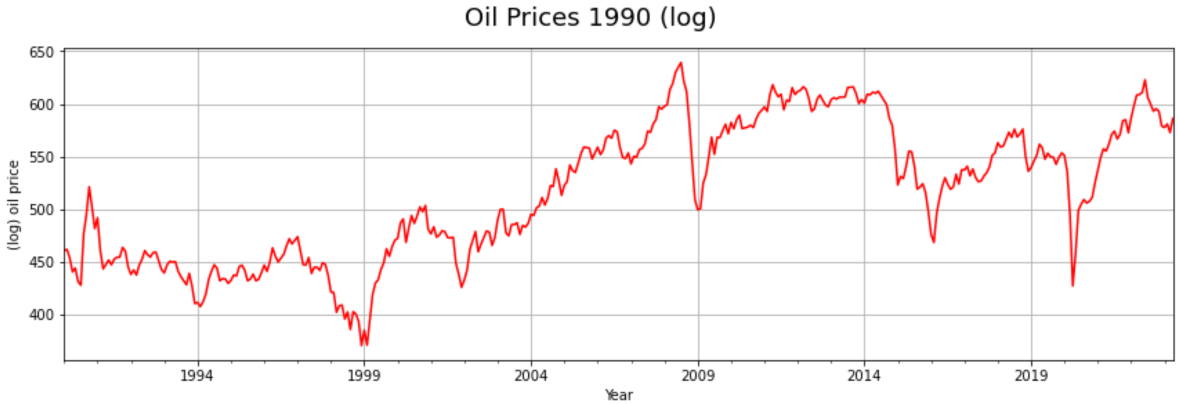


Figure 4. 2 Logarithm of the Oil Price Index

The graph above shows the producer price index of the oil price from January 1990 to April 2023 on a logarithmic scale. The logarithmic scales deflate the price level to accurately capture and visualize price changes (in percentage terms). As an example, after a logarithmic transformation, a given percentage change, say 5%, looks the same at any given time in the time series regardless of the price level at that given time.

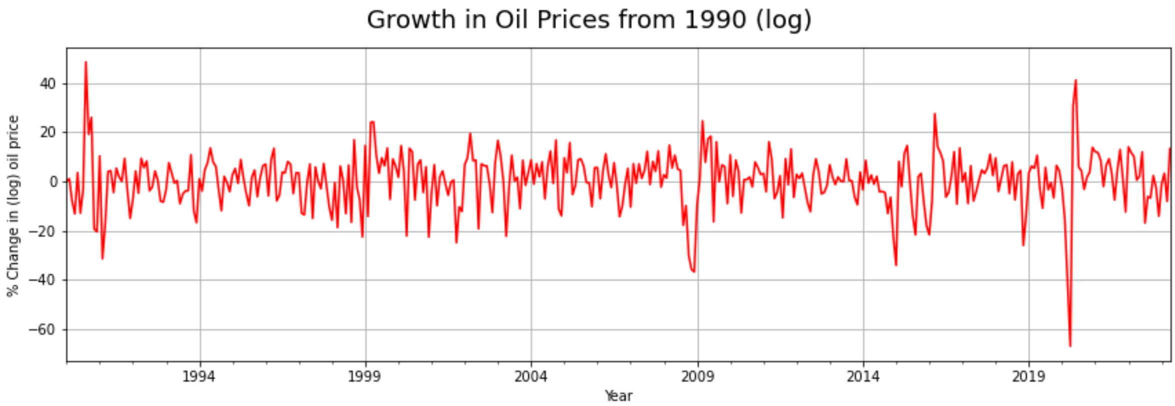


Figure 4. 3 Differentiation of the logarithm of the Oil Price Index

The graph above shows the producer price index of the oil price from January 1990 to April 2023. This graph is based on the logarithmic scale (in percentages) from above but further adjusted to look at the price change between any given period and its immediately preceding period, in other words, *period N – period N-1*.

4.1.2. Wheat Prices

As mentioned in subchapter *4.1. Data*, the data for the Global Wheat Price is retrieved from the Federal Reserve Bank of St. Louis, henceforth abbreviated to FRED, which specify their source for the time series in question to be the International Monetary Fund. The wheat price I use is named PWHEAMTUSDM and the values represent the benchmark prices representative of the global wheat market. It consists of monthly observations from January 1990, and is still published every month, as the most recent observation was published on the 11th of July 2023 for the month of June 2023 (International Monetary Fund, 2023). The unit of this time series is United States Dollar per metric ton of wheat. I have used this time series to create an index equaling 100 in January 1990 to ensure the same value at the starting point of the two time series I am analyzing. The data is not seasonally adjusted nor adjusted for inflation and consists of a total of 400 monthly observations from 1990-01-01:2023-04-01 (International Monetary Fund, 2023).

The International Monetary Fund describes the time series as the *No.1 Hard Red Winter Wheat, ordinary protein, Kansas City, US$ per metric ton* which is one type of wheat (which is described in chapter two Background) and the Geographic Coverage of this time series data is described as representative of the global wheat market, which is determined by the largest import markets of a given commodity (The International Monetary Fund, 2023). Based on the name of the time series and the description from the Federal Reserve Fund (International Monetary Fund, 2023) as well as the International Monetary Fund (The International Monetary Fund, 2023) we assume this time series is, in fact, representative for the global wheat prices.

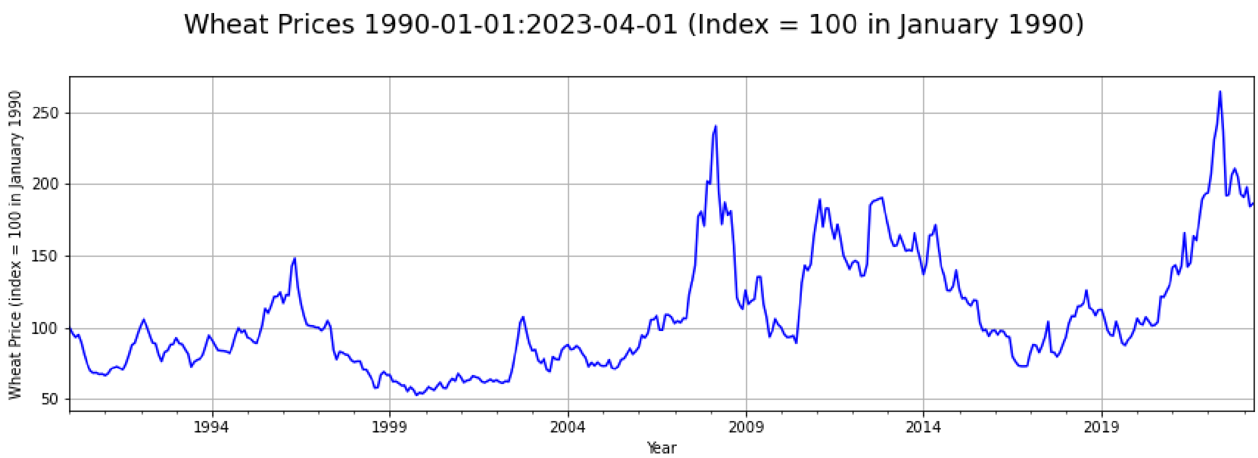


Figure 4. 4 Wheat Price Index

The graph above shows the price index of the global wheat price from January 1990 to April 2023 in nominal terms (remember the time series is indexed), based on the data from the International Monetary Fund.

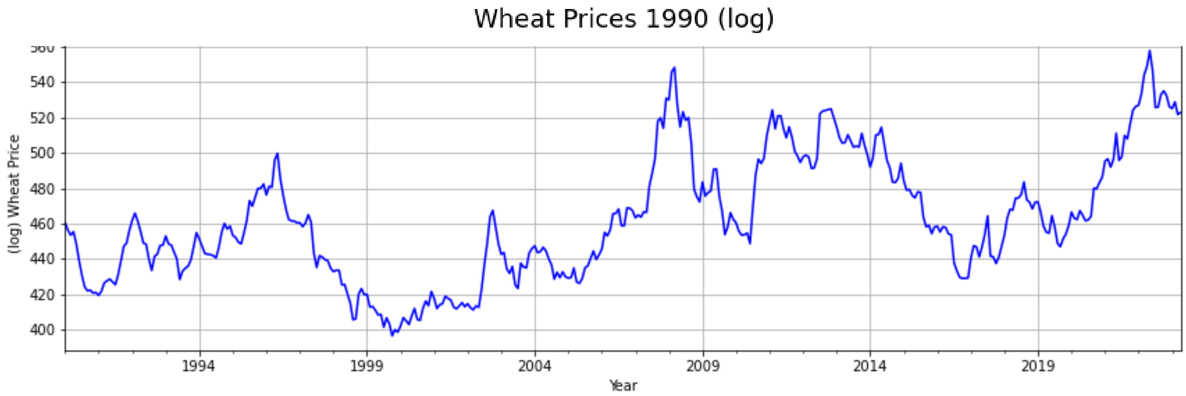


Figure 4. 5 Logarithm of the Wheat Price Index

The graph above shows the producer price index of the oil price from January 1990 to April 2023 on a logarithmic scale. The logarithmic scales deflate the price level to accurately capture and visualize price changes (in percentage terms). As an example, after a logarithmic transformation, a given percentage change, say 5%, looks the same at any given time in the time series regardless of the price level at that given time.

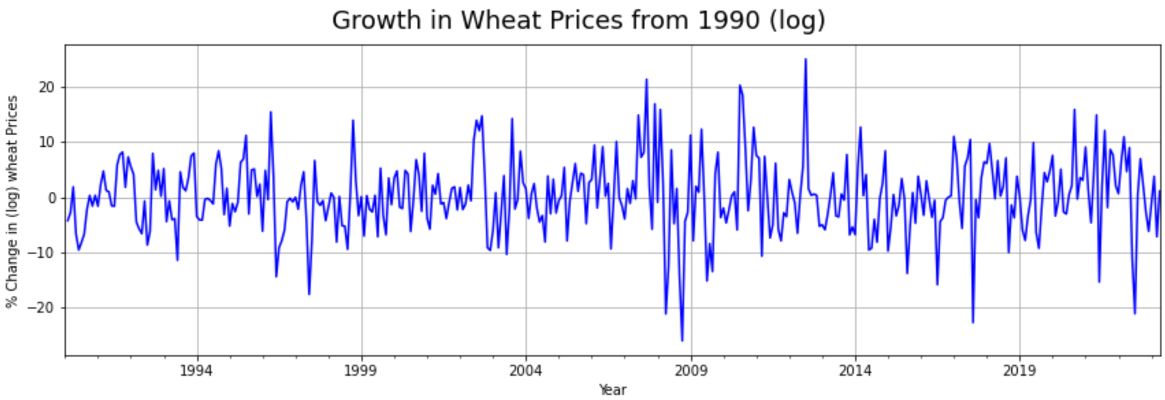


Figure 4. 6 Differentiation of the logarithm of the Wheat Price Index

The graph above shows the price index of the global wheat price from January 1990 to April 2023. This graph is based on the logarithmic scale (in percentages) from above but further adjusted to look at the price change between any given period and its immediately preceding period, in other words, *period N – period N-1*.

4.2. Methods

In this subchapter, we start by an overview of relevant literature to substantiate the method I use, then we continue by introducing you to autoregression model I use in my analysis, before I present key statistical concepts such as stationarity, autocorrelation and heteroskedasticity as well as econometric methods as the Augmented Dickey-Fuller test, the Built in Breusch-Godfrey test and Granger-Causality. Furthermore, hypothesis, parameters and the most relevant results from each statistical concept and method will be presented and explained.

Please note that, in our context, the wheat price index is the dependent variable (the endogenous variable), while the oil price index is the independent variable (the exogenous variable).

4.2.1. Literature review

This section presents a summary of pertinent literature relevant to my choice of model. I draw upon the findings of Hamilton (2003), who further expands on his analysis in Hamilton (2009) where he examines the relationship between oil prices and GDP. Employing his framework, I explore and assess the interconnectedness of the oil price and the wheat price.

Saghaian (2010) examines the interconnections between oil ethanol and corn in his paper. He uses monthly commodity price data from the Agricultural Statistics board (for wheat-, soybean- and corn prices per bushel), as well as crude oil- and ethanol prices per gallon from the Economic Research Service (USDA). His study period is from January 1996 to December 2008. He uses the method of differentiating before he tests for stationarity by using the Augmented Dickey-Fuller test (ADF test), then he performs a cointegration test to examine potential long-run relationships between the variables. He continues to specify a Vector Error Correction (VEC) model to conduct relevant hypothesis tests on his framework, he concludes by using the Directed Graph analysis and use a Granger-Causality test to examine the causality (on the Granger form) between the variables. He uses two lags in his Granger-Causality test. He concludes that, based on his VEC specification, there is a strong correlation, but no causal links between the energy- and the agricultural sector, in the sense that the energy system causes instability in the agricultural sector. He does find that the crude oil price does Granger-Cause the wheat-, soybean-, and corn prices Saghaian (2010).

Hamilton (2003) use quarterly data from 1947 to 2001 on quarterly growth rate on chained-weighted real GDP from the Bureau of Economic Analysis as well as 100 times the quarterly logarithmic growth rate of the nominal crude oil PPI (Producer Price Index) from three different sources, from 1947-1987 from Hamilton 1982, 1974-1999 from Citebase, UCDS library, and 1999-2001 the WPI0561 from the Bureau of Labor Statistics, respectively. As other more recent research suggests, he argues that there is a nonlinear relation between oil price changes and GDP growth. He bases this on the fact that high oil price growth typically reduces GDP growth but the same is not necessarily true in reverse. He also argues that an oil price increase, which happens in relatively close time proximity to an oil price decrease, in which restores the price to original levels, does not have the same effect as an increase after a longer stable period for the oil price. He uses this insight to create a nonlinear specification (which seems to be a variation of the ADL/ARDL method) based on capturing net oil price increases over an extended period. He defined this to be a three-year period (12 quarters) where the independent variable, $O_t^{\#}$, is set equal to zero if the observed oil price increase does not exceed any previous peak (in the t - 3 year period). However, if there is a price peak, said differently a 3 year high, then $O_t^{\#}$ represent the amount the oil price in period *t* exceed the value in the past 3 years Hamilton (2003).

Hamilton (2009) uses the nonlinear specification for his regression, described above Hamilton (2003), as well as a univariate autoregression model, or AR(p) model, where p indicates the number of lags, to analyze if oil prices change can predict the growth of GDP. The study period uses quarterly data starting in the second quarter of 1949 and spans to the fourth quarter of 2008. He uses the producer price index from the Bureau of Labor Statistics as the independent variable and the Real GDP provided by the U.S. Bureau of Economic Analysis as the dependent variable. In his paper, he compares his results with the re-estimated results on updated sample- and evaluation period to existing work which use different methods. This include Edelstein & Kilian (2007) use of AR(6) model on real PCE, PCE services, PCE nondurables, PCE durables, PCE autos and consumer sentiment, as well as VAR(5) model by Blanhard and Gali (2008) on real GDP. He use the univariate autoregression model as a baseline for comparing the improvements to the post-sample MSE (Mean Squared Error). His findings are inconclusive yet encouraging, as the result of his nonlinear specification model was greatly dependent upon the post-sample intervals. When using the same post-sample interval as (Edelstein & Kilian, 2007)and Blanhard and Gali (2008), his nonlinear specification model outperformed all the methods examined, resulting in a 45% improvement compared to the univariate autoregressive model with four lags Hamilton (2009).

4.2.2. Autoregression Model, AR(*p*)

The standard AR(*p*) model, short for the Autoregression Model of order *p*, where *p* represents the number of lags, is a linear univariate autoregression model. This implies we look at a one-way causal relationship in a linear model based on its own lags, otherwise known as the lags of the dependent variable, which Stock & Watson (2015) presents as:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + u_t \quad (1)$$

Where Y is a linear function, β_0 is the intercept, β_1, β_2,…, β_p are the parameters, *p* represent the number of lags, Y_{t-1}, Y_{t-2},…, Y_{t-p} represent the value of the lags and u_t is the error term.

Please note they assume: $E(u_t | Y_{t-1}, Y_{t-2},\ldots) = 0$. In other words, they assume strict exogeneity to comply with the OLS (Ordinary Least Squares) assumption Stock & Watson (2015).

The ADL(p,q) model, also known as the ARDL model, is short for the Autoregressive Distributed Lag Model and is a variation of the AR(*p*) model described above, where *p* represents the number of lags for Y_t and *q* the number of lags for X_t. In this case there is a dependent- (Y_t) and an independent (X_t) variable, where lags from both are used in predicting the future value of Y_t, which (Stock & Watson, 2015) present as:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \ldots + \delta_q X_{t-q} + u_t \quad (2)$$

Where Y is a linear function, β_0 is the intercept, β_1, β_2,…, β_p as well as δ_1, δ_2,…, δ_q are the parameters, *p* and *q*, respectively are the number of lags, Y_{t-1}, Y_{t-2},…, Y_{t-p} as well as X_{t-1}, X_{t-2},…, X_{t-q} represent the value of the lags and u_t is the error term.

As above they assume strict exogeneity: $E(u_t | Y_{t-1}, Y_{t-2},\ldots, X_{t-1}, X_{t-2},\ldots) = 0$.

However, (Hamilton, What is an oil shock?, 2003) develop a nonlinear specification for OLS, which I present below. Note that I have replaced his regression coefficients results with general coefficients, β and δ:

With four lags:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \delta_1 X^{\#}_{t-1} + \delta_2 X^{\#}_{t-2} + \delta_3 X^{\#}_{t-3} + \delta_4 X^{\#}_{t-4} \quad (3)$$

With an unspecified number of lags:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \delta_1 X^{\#}_{t-1} + \delta_2 X^{\#}_{t-2} + \ldots + \delta_q X^{\#}_{t-q} \quad (4)$$

Please note that this specification is linear in its parameters, but nonlinear in the X variable.

Furthermore, (Hamilton, Causes and Consequences of the Oil Shock of 2007-08, 2009) uses this nonlinear specification presented above, to compare the improvements in the MSE (Mean Squared Errors) to other models (as the VAR(5), AR(6) and AR(4)) with encouraging results. Please note that his specification seems to be a variation of the ADL/ARDL model which has a nonlinear independent variable.

I base my method on the nonlinear specification model derived by (Hamilton, What is an oil shock?, 2003). The specification is described, on the general form, below:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \delta_1 X^{\#}_{t-1} + \delta_2 X^{\#}_{t-2} + \ldots + \delta_q X^{\#}_{t-q} + u_t \quad (5)$$

Where Y is the dependent variable (wheat price change) and X is the independent variable (oil price change), where β is the parameter estimate for Y, and δ is the parameter estimate for X, t is the notation for time/period and # counts the variable if it is higher than any peak in the last year (12 months).

4.2.3. Test for stationarity and unit root by using the Augmented Dickey fuller test.

In time series regression and forecasting, stationarity is an important concept which refers to the idea that historic relationships in data can be generalized to forecast the future. For a time-series to be stationary, the probability distribution for a given variable in this time series must not change over time Stock & Watson (2015). In other words, stationarity requires that the joint distribution of a sequence (for example the first two observations in the time series in question) are the same as the joint distribution of any two observations in said time series or that “Stationarity does require that the nature of any correlation between adjacent terms is the same across all time periods” Wooldridge (2021).

In an autoregressive model, AR(p), where p represents the order (or number of lags), the model has a unit root and contains a stochastic trend, if the root equals 1. If the observations in the time series, at any given time is stationary, the time series does not have a unit root or a stochastic trend Stock & Watson (2015).

The hypothesis are as follows:

The null hypothesis (H0): Data has a unit root and is non-stationary.

The alternative hypothesis (HA): Data has no unit root and is stationary.

(Statology, (Dickey-Fuller Test), 2021)

As explained above, a stationary time series does not have a unit root and vice versa. We use the p-value yielded by this test statistic to reject or accept the null hypothesis. If the p-value is less than the significance level of 5% (<0,05) we reject the null hypothesis, if the p-value is greater than the significance level of 5% (>0,05) we accept the null hypothesis (Statology, Dickey-Fuller Test, 2021).

4.2.3.1. Detrending the data:

Our initial Dickey-Fuller test, presented above, reveals that our data has a unit root and is non-stationary. I use differencing to detrend, i.e., removing seasonal and or cyclical sub-trends from our data, to eliminate unit roots and get our data stationary for further analysis. The method of detrending consists of creating a new dataset where each new observation consists of the difference between any observation and the immediately preceding observation (Statology, 2021).

4.2.4. Test for Autocorrelation by using the Built in Breusch-Godfrey test.

We use the Breusch-Godfrey test in Python to test for autocorrelation of higher orders. The test we're using is the *acorr_breusch_godfrey()* function from the *statsmodels* library we imported. Remember we operate at a significance level of 5% (0,05) (Statology, Breusch-Godfrey Test, 2021).

The hypothesis are as follows:

The null hypothesis (H0): There is no autocorrelation at any order less than or equal to *p.*

The alternative hypothesis (HA): there exists autocorrelation at some order less than or equal to *p.*

(Statology, (Breusch-Godfrey Test), 2021)

We use the p-value yielded by this test statistic to reject or accept the null hypothesis. If the p-value is less than the significance level of 5% (<0,05) we reject the null hypothesis, if the p-value is greater than the significance level of 5% (>0,05) we accept the null hypothesis. Autocorrelation in the residuals should be addressed (Statology, Breusch-Godfrey Test, 2021).

4.2.5. Test for Heteroskedasticity (Breusch-Pagan test)

Heteroskedasticity is a term in regression analysis which refers to the unequal scatter of residuals, where the spread of the residuals changes over the time range of the measured values. Heteroskedasticity violates the OLS (Ordinary Least Squares) assumption of homoskedasticity which implies constant variance, and the results are hard to trust if heteroskedasticity exists. The Breusch-Pagan test is used to determine if heteroskedasticity is present in our regression analysis (Statology, Breusch-Pagan Test, 2020).

The hypothesis are as follows:

The null hypothesis (H0): homoskedasticity is present.

The alternative hypothesis (HA): homoskedasticity is not present.

(Statology, Breusch-Pagan Test, 2020)

We use the p-value yielded by this test statistic to reject or accept the null hypothesis. If the p-value is less than the significance level of 5% (<0,05) we reject the null hypothesis, if the p-value is greater than the significance level of 5% (>0,05) we accept the null hypothesis. Heteroskedasticity, the unequal scatter of residuals, should be addressed (Statology, Breusch-Pagan Test, 2020).

4.2.6. Test for Granger-Causality (of oil price change)

In econometrics *causality* is usually defined as an observed effect on Y based on different values of X, ideally in the context of a randomized controlled experiment. In other words, there is causality when a change in X is the direct cause for the change observed in Y. Granger-Causality can be misleading as it does not refer to causality as defined above, if X Granger-Causes Y, the correct interpretation is that X is useful in predicting Y. Stated differently, it means that historical X observations has information in its observations which is useful to predict Y, beyond the historical observations of Y itself (Stock & Watson, 2015).

I use the F-Statistic (from an F-test in Python) to test if the oil price Granger-Causes the wheat price. A significant F-Statistic indicates the regression model makes a better fit for the data compared to a regression model without any independent variables (remember the oil price index is the independent variable in this context) and vice versa. The F-test reports the F-statistic as well as its corresponding p-value, if the p-value is lower than our significance level (the significance level is 5% or 0,05 in this analysis), we conclude this regression model is a better fit for the data compared to the model without any predictor variables (Statology, F-test, 2019).

The hypothesis are as follows:

The null hypothesis (H0): Time series for oil price does not Granger-Cause time series Wheat.

The alternative hypothesis (HA): Time series for oil price Granger-Cause time series Wheat.

We use the p-value corresponding to the F-Statistic yielded by the F-test to reject or accept the null hypothesis. If the p-value is less than the significance level of 5% (<0,05) we reject the null hypothesis, if the p-value is greater than the significance level of 5% (>0,05) we accept the null hypothesis (Statology, F-test, 2019).

4.2.7. AIC and BIC to aid in determining the number of lags

I use the AIC (Akaike Information Criterion) and BIC (Bayes Information Criterion) results from our models to help determine the number of lags. The BIC results are especially helpful in determining the number of lags to include in the regression model, as the result with the lowest value is considered to be the consistent estimator which presents the true lag length, if the BIC is not available, the AIC could be used instead (Stock & Watson, 2015).

5. Results

This chapter is divided into five main subchapters, each subchapter is based on a stated time interval from our data, spanning from January 1990 to April 2023. Please note that the intervals are constructed to include or exclude events, shocks, and changes which might influence the results of the model. This allows for a comprehensive comparison and analysis of the model's performance under varying circumstances.

To clearly distinguish between the time intervals, each interval is labeled A, B, C, D or E and is presented in separate subchapters below. Time interval A spans from January 1990 to January 2021, time interval B spans from January 1990 to January 2020, time interval C spans from January 1990 to April 2023, time interval D spans from January 1990 to January 2010 and time interval E spans from January 2010 to April 2023.

The selection of time intervals A and B aimed to examine whether the inclusion of the initial year of the covid-19 pandemic would impact the findings of my analysis. Time interval C was selected to encompass Russia's military invasion of Ukraine and the period until most travel and lockdown restrictions were lifted worldwide due to covid-19. Time intervals D and E were chosen to explore any disparities before and after the rise in popularity of biofuel, as discussed in chapter 3, as well as before and after the global financial crisis of 2007-2009, as discussed in chapter 2.

In chapter four, I present two distinct models to analyze the data. The first model is a standard univariate autoregressive model, whereas the second model is a nonlinear specification model that builds upon the framework by Hamilton discussed in chapter four.

The results section for each time interval is divided into three main sections. The first section provides a summary of the AIC and BIC scores for both models, whereas the second section presents the granger-causality results based on the models, considering the five different lag lengths which I test for. The third section focuses on forecasting, in which I generate up to six different forecasts for each time interval. I depend on the BIC outcomes and the associated p-values of the F-statistic from the models to ascertain the number of lags I will employ for each forecast. Please note that I solely evaluate the models for 3-, 4-, 5-, 6-, and 8-lags, therefore it is important to acknowledge that if the BIC indicates that the optimal number of lags lies at either extreme, the actual optimal number of lags may be lower or higher than the one I selected for my forecast analysis.

5.1. Results for time interval A

Time interval A starts in January 1990 and ends in January 2021. This time series includes the first year of the global pandemic, and the stringent lockdowns of countries across the globe.

5.1.1. Summary of the AIC and BIC tests on the models with different lags

Table 5. 1 AIC and BIC results based on the number of lags for each model.

AIC and BIC Results dependent of number of lags	AIC	BIC
Nonlinear specification (ARDL/ADL) Model - 3 lags	2425,38	**2452,76**
Nonlinear specification (ARDL/ADL) Model - 4 lags	2420,42	2455,59
Nonlinear specification (ARDL/ADL) Model - 5 lags	2412,52	2455,48
Nonlinear specification (ARDL/ADL) Model - 6 lags	2409,87	2460,60
Nonlinear specification (ARDL/ADL) Model - 8 lags	**2397,13**	2463,39
Univariate Autoregression Model - 3 lags	2427,60	2443,24
Univariate Autoregression Model - 4 lags	2422,70	2442,25
Univariate Autoregression Model - 5 lags	2415,03	2438,46
Univariate Autoregression Model - 6 lags	2410,36	2437,68
Univariate Autoregression Model - 8 lags	**2396,93**	**2432,01**

The results of the AIC and BIC for both models with different number of lags are presented in the table above. Please note that the models with the lowest AIC or BIC number are highlighted in bold and are considered the best lag length choice for each model.

5.1.2. Summary of Granger-Causality results

In this subchapter the F-statistic as well as its corresponding P-value from the Granger-Causality Tests are presented in two different tables, the first for the nonlinear specification model, the second for the univariate autoregression model. Please note that the results are presented for all the lag lengths I test for, and that N/A indicate that at least one statistical test failed at the 5% significance level.

Table 5. 2 Results from the Granger-Causality test on the nonlinear specification model.

Results of Granger-Causality tests with different lags

Null Hypothesis - nonlinear specification model (ARDL/ADL)	F-statistic	Corresponding P-value
Oil price does not Granger-Cause wheat prices 1990-2023, 3 lags	3,0091	0,03025**
Oil price does not Granger-Cause wheat prices 1990-2023, 4 lags	2,8330	0,02453**
Oil price does not Granger-Cause wheat prices 1990-2023, 5 lags	2,5120	0,02976**
Oil price does not Granger-Cause wheat prices 1990-2023, 6 lags	2,0677	0,05637*
Oil price does not Granger-Cause wheat prices 1990-2023, 8 lags	1,9089	0,05773*

*** Significant at the 1% level; **Significant at the 5% level, *Significant at the 10% level

Table 5. 3 Results from the Granger-Causality test on the univariate autoregression model.

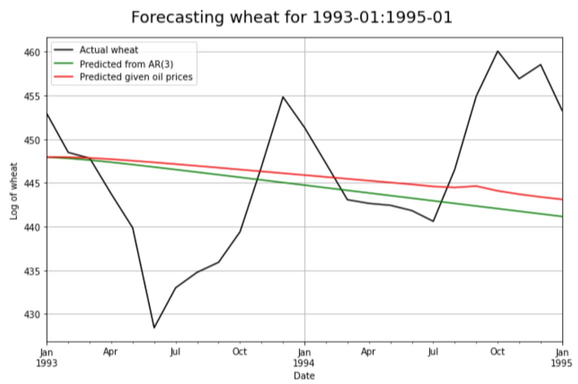
Results of Granger-Causality tests with different lags

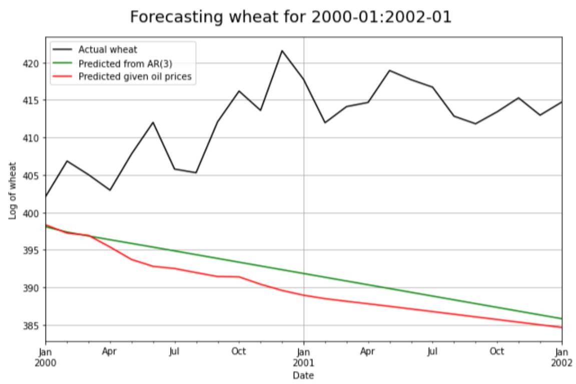
Null Hypothesis - Univariate autoregression model	F-statistic	Corresponding P-value
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 3 lags	7,5053	0,00007***
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 4 lags	6,2025	0,00008***
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 5 lags	4,7319	0,00034***
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 6 lags	3,7939	0,00112***
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 8 lags	N/A	N/A

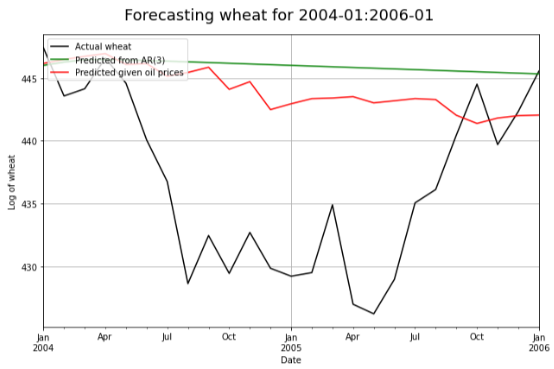
*** Significant at the 1% level; **Significant at the 5% level, *Significant at the 10% level

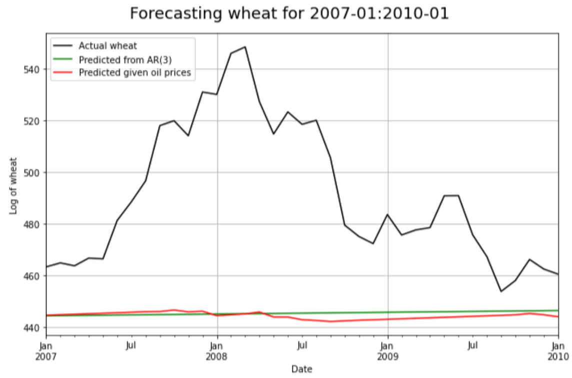
5.1.3. Forecasting for time interval A

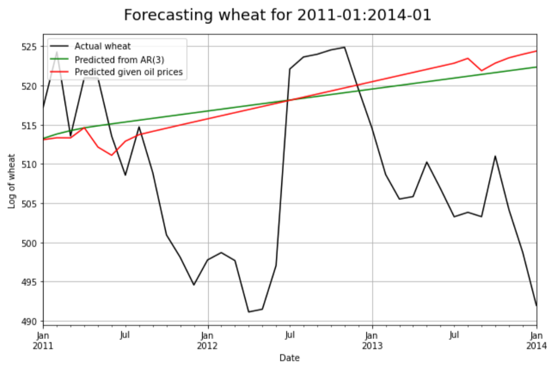
In this subchapter I am going to present six forecasts for time interval A, which spans from January 1990 to January 2021, for the nonlinear specification model inspired by Hamilton. Out of the six forecasts, three of them represent a time interval the United States is in a recession, while the remaining three when the United States is not in a recession. The lag selection is based on the BIC score for each of the five lag lengths I test for, presented in Table 5.1. above, which is three lags for time interval A. The forecast in this subchapter is therefore based on three lags for both models. Please note that the forecasting is based on the optimal lag selection from the nonlinear specification model only, in other words, there will not be performed or presented the parallel results based on the optimal lag selection for the univariate autoregression model, which is eight lags for this time interval.

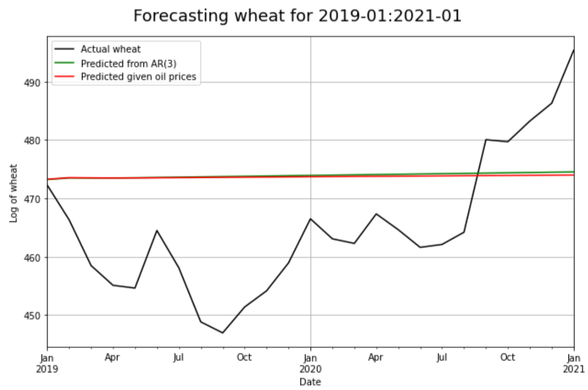





Forecasting wheat for 2004-01:2006-01
Actual wheat
Predicted from AR(3)
Predicted given oil prices
Log of wheat
Date


Forecasting wheat for 2007-01:2010-01
Actual wheat
Predicted from AR(3)
Predicted given oil prices
Log of wheat
Date


Forecasting wheat for 2011-01:2014-01
Actual wheat
Predicted from AR(3)
Predicted given oil prices
Log of wheat
Date


Forecasting wheat for 2019-01:2021-01
Actual wheat
Predicted from AR(3)
Predicted given oil prices
Log of wheat
Date

5.2. Results for time interval B

Time interval B starts in January 1990 and ends in January 2020. Time series end before Covid-19 and Russia's invasion of Ukraine, two events which significantly disrupts global trade in general, and oil and wheat in particular.

5.2.1. Summary of the AIC and BIC tests on the models with different lags

Table 5. 4 AIC and BIC results based on the number of lags for each model.

AIC and BIC Results dependent of number of lags	AIC	BIC
Nonlinear specification (ARDL/ADL) Model - 3 lags	2348,8	**2375,94**
Nonlinear specification (ARDL/ADL) Model - 4 lags	2344,05	2378,92
Nonlinear specification (ARDL/ADL) Model - 5 lags	2336,62	2379,21
Nonlinear specification (ARDL/ADL) Model - 6 lags	2333,98	2384,28
Nonlinear specification (ARDL/ADL) Model - 8 lags	**2321,94**	2387,62
Univariate Autoregression Model - 3 lags	2350,61	2366,12
Univariate Autoregression Model - 4 lags	2345,71	2365,08
Univariate Autoregression Model - 5 lags	2338,27	2361,5
Univariate Autoregression Model - 6 lags	2333,62	2360,7
Univariate Autoregression Model - 8 lags	**2320,27**	**2355,05**

The results of the AIC and BIC for both models with different number of lags are presented in the table above. Please note that the models with the lowest AIC or BIC number is highlighted in bold and is considered the best lag length choice for each model.

5.2.2. Summary of Granger-Causality results

In this subchapter the F-statistic as well as its corresponding P-value from the Granger-Causality Tests are presented in two different tables, the first for the nonlinear specification model, the second for the univariate autoregression model. Please note that the results are presented for all the lag lengths I test for, and that N/A indicate that at least one statistical test failed at the 5% significance level.

Table 5. 5 Results from the Granger-Causality test on the nonlinear specification model.

Results of Granger-Causality tests with different lags

Null Hypothesis - nonlinear specification model (ARDL/ADL)	F-statistic	Corresponding P-value
Oil price does not Granger-Cause wheat prices 1990-2020, 3 lags	**2,8402**	0,03790**
Oil price does not Granger-Cause wheat prices 1990-2020, 4 lags	2,6778	**0,03172****
Oil price does not Granger-Cause wheat prices 1990-2020, 5 lags	2,3234	0,04274**
Oil price does not Granger-Cause wheat prices 1990-2020, 6 lags	1,9054	0,07930*
Oil price does not Granger-Cause wheat prices 1990-2020, 8 lags	1,6977	0,09872*

*** Significant at the 1% level; **Significant at the 5% level, *Significant at the 10% level

Table 5. 6 Results from the Granger-Causality test on the univariate autoregression model.

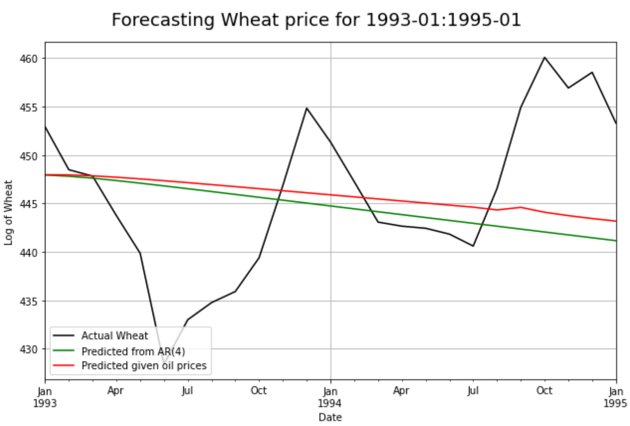
Results of Granger-Causality tests with different lags

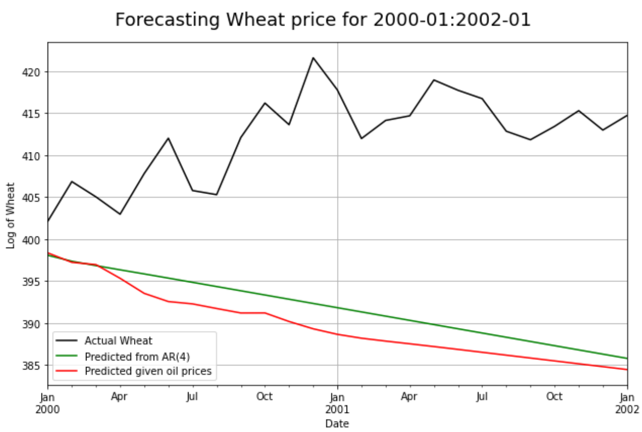
Null Hypothesis - Univariate autoregression model	F-statistic	Corresponding P-value
Previous wheat prices does not Granger-Cause wheat prices 1990-2020, 3 lags	**7,442**	**0.00001***
Previous wheat prices does not Granger-Cause wheat prices 1990-2020, 4 lags	6,0813	0.00010***
Previous wheat prices does not Granger-Cause wheat prices 1990-2020, 5 lags	4,6541	0,00040***
Previous wheat prices does not Granger-Cause wheat prices 1990-2020, 6 lags	3,7298	0,00132***
Previous wheat prices does not Granger-Cause wheat prices 1990-2020, 8 lags	N/A	N/A

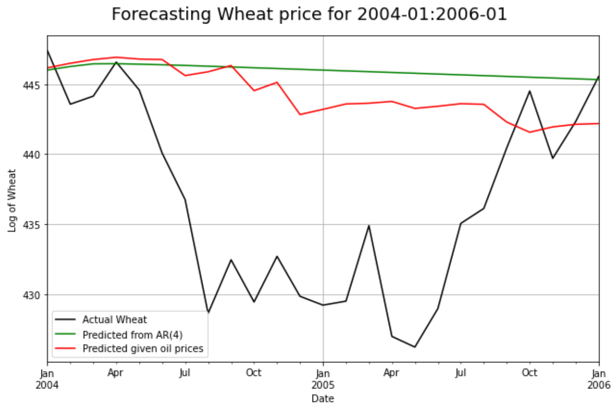
*** Significant at the 1% level; **Significant at the 5% level, *Significant at the 10% level

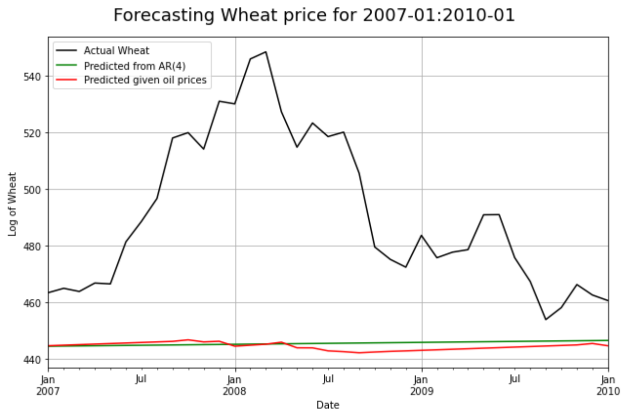
5.2.3. Forecasting for time interval B

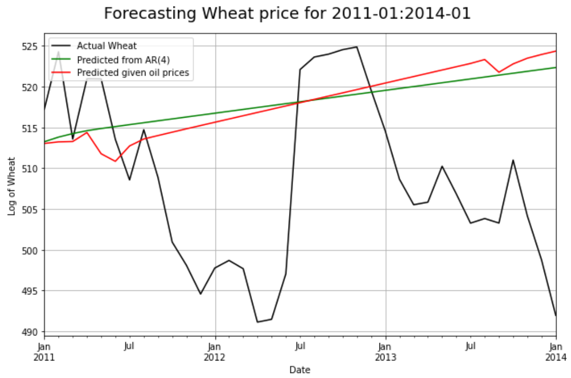
In this subchapter I am going to present five forecasts for time interval B, which spans from January 1990 to January 2020, for the nonlinear specification model inspired by Hamilton. Out of the five forecasts, two of which from a time interval the United States is in a recession, the remaining three when the United States is not in a recession. The lag selection is based on the BIC score for each of the five lag lengths I test for, presented in Table 5.4. above, which is three lags for time interval B. The forecast in this subchapter is therefore based on three lags for both models. Please note that the forecasting is based on the optimal lag selection from the nonlinear specification model only, in other words, there will not be performed or presented the parallel results based on the optimal lag selection for the univariate autoregression model, which is eight lags for this time interval.






Forecasting Wheat price for 2004-01:2006-01
Actual Wheat
Predicted from AR(4)
Predicted given oil prices
Log of Wheat
Date


Forecasting Wheat price for 2007-01:2010-01
Actual Wheat
Predicted from AR(4)
Predicted given oil prices
Log of Wheat
Date


Forecasting Wheat price for 2011-01:2014-01
Actual Wheat
Predicted from AR(4)
Predicted given oil prices
Log of Wheat
Date

5.3. Results for time interval C

Time interval C starts in January 1990 and ends in April 2023, it represents the entire time series I operate with for both commodities.

5.3.1. Summary of the AIC and BIC tests on the models with different lags

Table 5. 7 AIC and BIC results based on the number of lags for each model.

AIC and BIC Results dependent of number of lags	AIC	BIC
Nonlinear specification (ARDL/ADL) Model - 3 lags	2623,15	2651,02
Nonlinear specification (ARDL/ADL) Model - 4 lags	2614,58	**2650,39**
Nonlinear specification (ARDL/ADL) Model - 5 lags	2608,24	2651,98
Nonlinear specification (ARDL/ADL) Model - 6 lags	2605,59	2657,25
Nonlinear specification (ARDL/ADL) Model - 8 lags	**2595,27**	2662,74
Univariate Autoregression Model - 3 lags	2620,30	2636,23
Univariate Autoregression Model - 4 lags	2615,35	2635,24
Univariate Autoregression Model - 5 lags	2607,65	2631,51
Univariate Autoregression Model - 6 lags	2603,08	2630,90
Univariate Autoregression Model - 8 lags	**2590,55**	**2626,26**

The results of the AIC and BIC for both models with different number of lags are presented in the table above. Please note that the models with the lowest AIC or BIC number are highlighted in bold and are considered the best lag length choice for each model.

5.3.2. Summary of Granger-Causality results

In this subchapter the F-statistic as well as its corresponding P-value from the Granger-Causality Tests are presented in two different tables, the first for the nonlinear specification model, the second for the univariate autoregression model. Please note that the results are presented for all the lag lengths I test for, and that N/A indicate that at least one statistical test failed at the 5% significance level.

Table 5. 8 Results from the Granger-Causality test on the nonlinear specification model.

Results of Granger-Causality tests with different lags

Null Hypothesis - nonlinear specification model (ARDL/ADL)	F-statistic	Corresponding P-value
Oil price does not Granger-Cause wheat prices 1990-2023, 3 lags	N/A	N/A
Oil price does not Granger-Cause wheat prices 1990-2023, 4 lags	N/A	N/A
Oil price does not Granger-Cause wheat prices 1990-2023, 5 lags	1,6383	0,14885
Oil price does not Granger-Cause wheat prices 1990-2023, 6 lags	1,3992	0,21373
Oil price does not Granger-Cause wheat prices 1990-2023, 8 lags	1,1729	0,31431

*** Significant at the 1% level; **Significant at the 5% level, *Significant at the 10% level

Table 5. 9 Results from the Granger-Causality test on the univariate autoregression model.

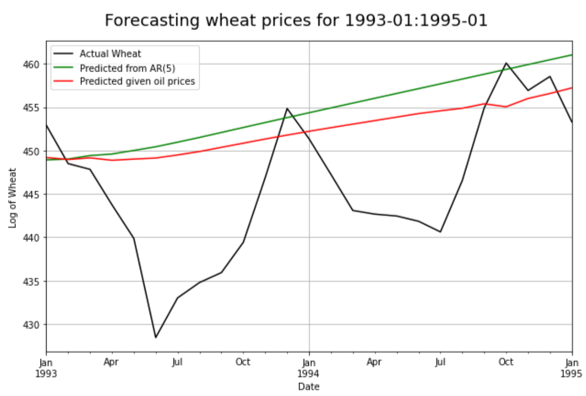
Results of Granger-Causality tests with different lags

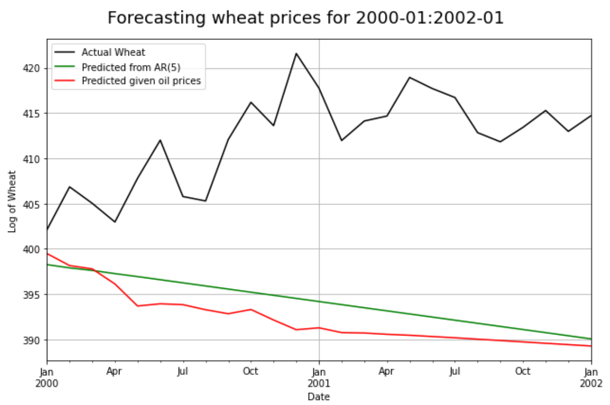
Null Hypothesis - Univariate autoregression model	F-statistic	Corresponding P-value
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 3 lags	5,7772	0,00071***
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 4 lags	4,8179	0,00084***
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 5 lags	3,8700	0,00196***
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 6 lags	3,1216	0,00535***
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 8 lags	N/A	N/A

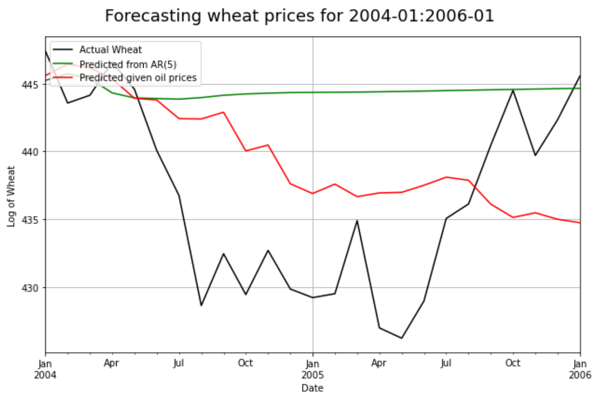
*** Significant at the 1% level; **Significant at the 5% level, *Significant at the 10% level

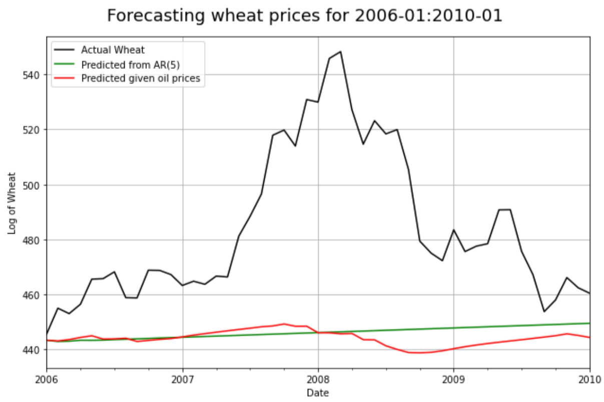
5.3.3. Forecasting for time interval C

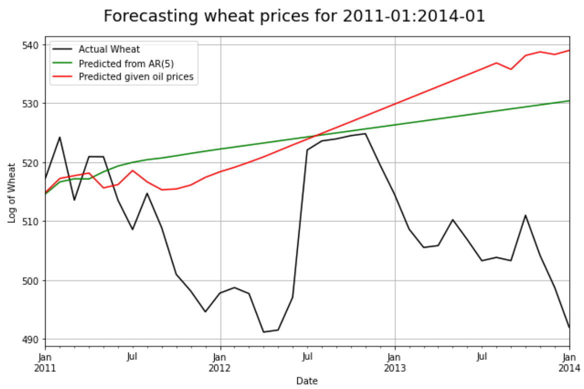
In this subchapter I am going to present six forecasts for time interval C, which spans from January 1990 to April 2023, for the nonlinear specification model inspired by Hamilton. Out of the six forecasts, three of them represent a time interval the United States is in a recession, while the remaining three when the United States is not in a recession. The lag selection is based on the BIC score for each of the five lag lengths I test for, presented in Table 5.7., which is four lags for time interval C. Due to the fact that the F-statistic and its corresponding p-value for the granger-causality test results are N/A for four lags, I will use five lags as it provides the greatest results in terms of the F-statistic and p-value whilst minimizing the BIC score given our constraint of a 5% significance interval for our statistic tests. The forecast in this subchapter is therefore based on five lags for both models. Please note that the forecasting is based on the optimal lag selection from the nonlinear specification model only, in other words, there will not be performed or presented the parallel results based on the optimal lag selection for the univariate autoregression model, which is eight lags for this time interval.

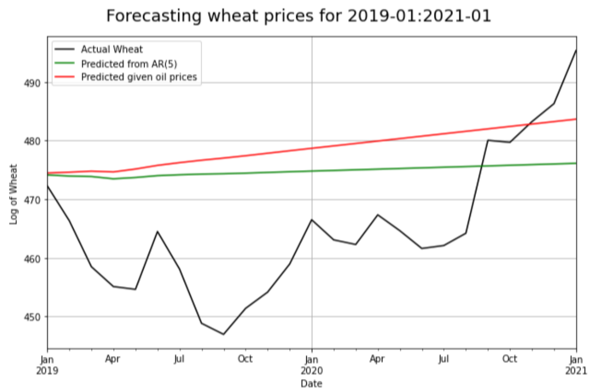





Forecasting wheat prices for 2004-01:2006-01
Actual Wheat
Predicted from AR(5)
Predicted given oil prices
Log of Wheat
Date


Forecasting wheat prices for 2006-01:2010-01
Actual Wheat
Predicted from AR(5)
Predicted given oil prices
Log of Wheat
Date


Forecasting wheat prices for 2011-01:2014-01
Actual Wheat
Predicted from AR(5)
Predicted given oil prices
Log of Wheat
Date


Forecasting wheat prices for 2019-01:2021-01
Actual Wheat
Predicted from AR(5)
Predicted given oil prices
Log of Wheat
Date

5.4. Results for time interval D

Time interval D starts in January 1990 and ends in January 2010. This time interval includes the financial crisis of 2008 but does not include events as the oil price crash in 2014, the global pandemic or Russia's invasion of Crimea peninsula (2014) or Ukraine (2022).

5.4.1. Summary of the AIC and BIC tests on the models with different lags

Table 5. 10 AIC and BIC results based on the number of lags for each model.

AIC and BIC Results dependent of number of lags	AIC	BIC
Nonlinear specification (ARDL/ADL) Model - 3 lags	1552,36	1576,64
Nonlinear specification (ARDL/ADL) Model - 4 lags	1544,79	1575,97
Nonlinear specification (ARDL/ADL) Model - 5 lags	1536,73	**1574,79**
Nonlinear specification (ARDL/ADL) Model - 6 lags	1533,56	1578,48
Nonlinear specification (ARDL/ADL) Model - 8 lags	**1520,93**	1579,53
Univariate Autoregression Model - 3 lags	1552,02	1565,89
Univariate Autoregression Model - 4 lags	1546,31	1563,63
Univariate Autoregression Model - 5 lags	1538,99	1559,74
Univariate Autoregression Model - 6 lags	1534,49	1558,68
Univariate Autoregression Model - 8 lags	**1520,98**	**1552,00**

The results of the AIC and BIC for both models with different number of lags are presented in the table above. Please note that the models with the lowest AIC or BIC number are highlighted in bold and are considered the best lag length choice for each model.

5.4.2. Summary of Granger-Causality results

In this subchapter the F-statistic as well as its corresponding P-value from the Granger-Causality Tests are presented in two different tables, the first for the nonlinear specification model, the second for the univariate autoregression model. Please note that the results are presented for all the lag lengths I test for, and that N/A indicate that at least one statistical test failed at the 5% significance level.

Table 5. 11 Results from the Granger-Causality test on the nonlinear specification model.

Results of Granger-Causality tests with different lags

Null Hypothesis - nonlinear specification model (ARDL/ADL)	F-statistic	Corresponding P-value
Oil price does not Granger-Cause wheat prices 1990-2023, 3 lags	N/A	N/A
Oil price does not Granger-Cause wheat prices 1990-2023, 4 lags	2,4904	0,04408**
Oil price does not Granger-Cause wheat prices 1990-2023, 5 lags	2,0655	0,07072*
Oil price does not Granger-Cause wheat prices 1990-2023, 6 lags	1,9737	0,07051*
Oil price does not Granger-Cause wheat prices 1990-2023, 8 lags	2,0013	0,04759**

*** Significant at the 1% level; **Significant at the 5% level, *Significant at the 10% level

Table 5. 12 Results from the Granger-Causality test on the univariate autoregression model.

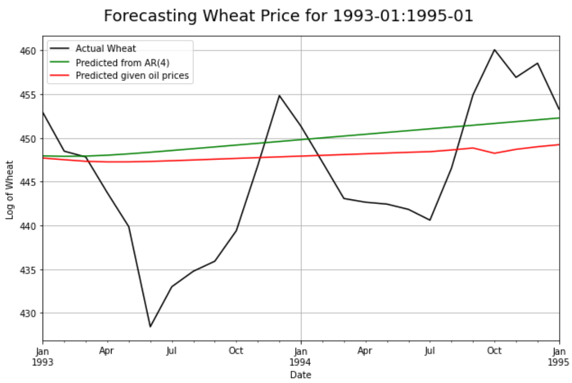
Results of Granger-Causality tests with different lags

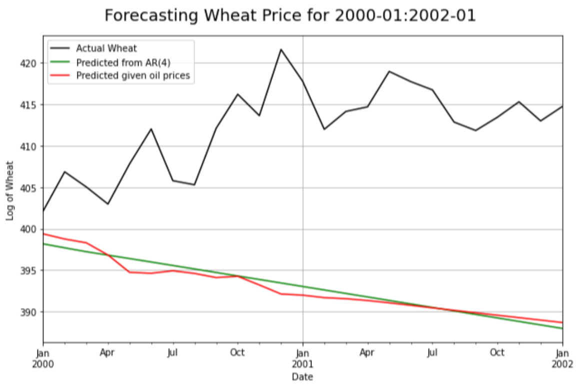
Null Hypothesis - Univariate autoregression model	F-statistic	Corresponding P-value
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 3 lags	5,6500	0,00094***
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 4 lags	4,6923	0,00117***
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 5 lags	3,3818	0,00577***
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 6 lags	2,7347	0,01392**
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 8 lags	2,2828	0,02302**

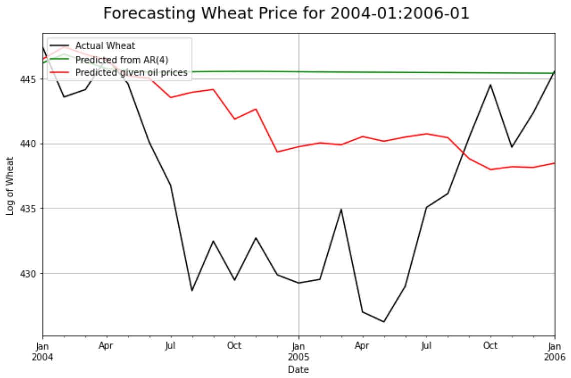
*** Significant at the 1% level; **Significant at the 5% level, *Significant at the 10% level

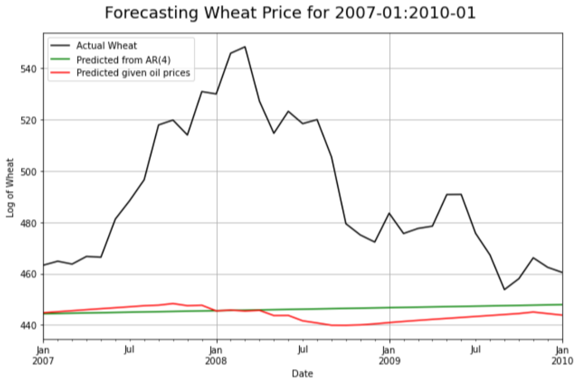
5.4.3. Forecasting for time interval D

In this subchapter I am going to present four forecasts for time interval D, which spans from January 1990 to January 2010, for the nonlinear specification model inspired by Hamilton. Out of the four forecasts, two of them represent a time interval the United States is in a recession, while the remaining two when the United States is not in a recession. The lag selection is based on the BIC score for each of the five lag lengths I test for, presented in Table 5.10., which is five lags for time interval D. Due to the fact that the F-statistic and its corresponding p-value for the granger-causality test results are not significant at the 5% level for five lags, but are significant for four lags, I will use four lags while forecasting time interval D. The forecast in this subchapter is therefore based on four lags for both models, in other words, there will not be performed or presented the parallel results based on the optimal lag selection for the univariate autoregression model, which is eight lags for this time interval.






Forecasting Wheat Price for 2004-01:2006-01
Actual Wheat
Predicted from AR(4)
Predicted given oil prices
Log of Wheat
Date


Forecasting Wheat Price for 2007-01:2010-01
Actual Wheat
Predicted from AR(4)
Predicted given oil prices
Log of Wheat
Date

5.5. Results for time interval E

Time interval E starts in January 2010 and ends in April 2023.

5.5.1. Summary of the AIC and BIC tests on the models with different lags

Table 5. 13 AIC and BIC results based on the number of lags for each model.

AIC and BIC Results dependent of number of lags	AIC	BIC
Nonlinear specification (ARDL/ADL) Model - 3 lags	1081,29	1102,82
Nonlinear specification (ARDL/ADL) Model - 4 lags	1083,04	1110,71
Nonlinear specification (ARDL/ADL) Model - 5 lags	**1067,76**	**1101,38**
Nonlinear specification (ARDL/ADL) Model - 6 lags	1089,62	1129,59
Nonlinear specification (ARDL/ADL) Model - 8 lags	1094,41	1146,69
Univariate Autoregression Model - 3 lags	1078,93	1091,23
Univariate Autoregression Model - 4 lags	1080,70	1096,08
Univariate Autoregression Model - 5 lags	1063,07	1081,41
Univariate Autoregression Model - 6 lags	1083,95	1105,47
Univariate Autoregression Model - 8 lags	1087,63	1115,31

The results of the AIC and BIC for both models with different number of lags are presented in the table above. Please note that the models with the lowest AIC or BIC number are highlighted in bold and are considered the best lag length choice for each model.

5.5.2. Summary of Granger-Causality results

In this subchapter the F-statistic as well as its corresponding P-value from the Granger-Causality Tests are presented in two different tables, the first for the nonlinear specification model, the second for the univariate autoregression model. Please note that the results are presented for all the lag lengths I test for, and that N/A indicate that at least one statistical test failed at the 5% significance level.

Table 5. 14 Results from the Granger-Causality test on the nonlinear specification model.

Results of Granger-Causality tests with different lags

Null Hypothesis - nonlinear specification model (ARDL/ADL)	F-statistic	Corresponding P-value
Oil price does not Granger-Cause wheat prices 1990-2023, 3 lags	0,9812	0,40332
Oil price does not Granger-Cause wheat prices 1990-2023, 4 lags	1,2108	0,30859
Oil price does not Granger-Cause wheat prices 1990-2023, 5 lags	0,8350	0,52684
Oil price does not Granger-Cause wheat prices 1990-2023, 6 lags	0,7249	0,63020
Oil price does not Granger-Cause wheat prices 1990-2023, 8 lags	0,8164	0,58950

*** Significant at the 1% level; **Significant at the 5% level, *Significant at the 10% level

Table 5. 15 Results from the Granger-Causality test on the univariate autoregression model.

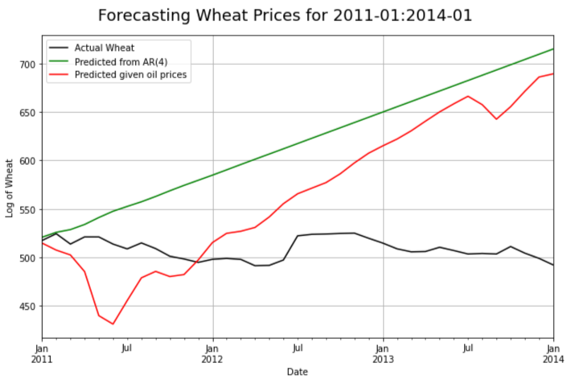
Results of Granger-Causality tests with different lags

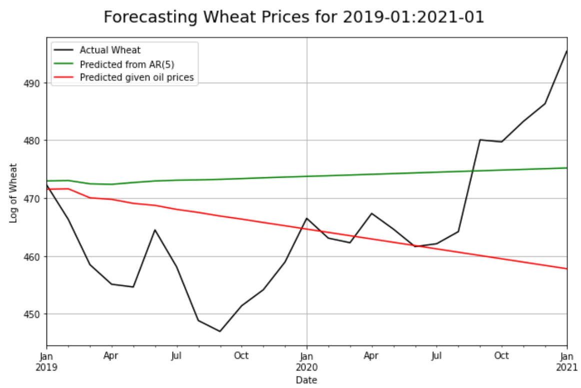
Null Hypothesis - Univariate autoregression model	F-statistic	Corresponding P-value
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 3 lags	1,7959	0,15030
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 4 lags	1,3096	0,26900
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 5 lags	1,1613	0,33107
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 6 lags	0,9927	0,43248
Previous wheat prices does not Granger-Cause wheat prices 1990-2023, 8 lags	0,9338	0,49073

*** Significant at the 1% level; **Significant at the 5% level, *Significant at the 10% level

5.5.3. Forecasting for time interval E

In this subchapter I am going to present two forecasts for time interval E, which spans from January 2010 to April 2023, for the nonlinear specification model inspired by Hamilton. Out of the two forecasts, one of them represent a time interval the United States is in a recession, while the remaining is when the United States is not in a recession. The lag selection is based on the BIC score for each of the five lag lengths I test for, presented in Table 5.13., which is five lags for time interval E. The forecast in this subchapter is therefore based on five lags for both models. The forecasting is based on the optimal lag selection from the nonlinear specification model only, in other words, there will not be performed or presented the parallel results based on the optimal lag selection for the univariate autoregression model, which is eight lags for this time interval. Please note that none of the lag lengths for this time interval has Granger-Causality for either model at the 5% significance level.





6. Concluding remarks and recommendations

This chapter is divided into three subchapters. The initial subchapter provides a concise overview of the discoveries made in chapter five. The second subchapter focuses on the limitations of this study, while the third subchapter provides recommendations and suggestions for further research.

6.1. Summary of findings

To summarize the findings for the nonlinear specification model: for time interval A, B and D there is at least one number of lags with provide statistical significance at the 5% level. For time interval C and E there are zero number of lags which provides statistical significance at the 5% level. To summarize the finding for the univariate autoregression model: for time interval A, B, C and D there is at least one number of lags with statistical significance at the 5% level. For time interval E there are zero number of lags which provides statistical significance at the 5% level. I conclude that this model is more likely to result in oil price granger-causing the wheat price for time intervals that end no later than January 2021.

Based on my finding, for time interval A, B and D I reject my null hypothesis and accept my alternative hypothesis that the oil price granger-causes the wheat price for both models. I additionally reject the null hypothesis and accepts the alternative hypothesis for the univariate autoregression model for time interval C. I accept my null hypothesis for time interval E for both models, I additionally accept my null hypothesis for time interval C for the nonlinear specification model.

The finding of my research is consistent with the recent study conducted by Umar et. al. (2021). This study concluded that an oil shock has a significant granger-causality for grains, live cattle, and wheat. It is important to note that the statistical significance is only observed at the 5% level for supply shocks, not demand shocks, between the oil and wheat price in their study. These findings are in line with my results as I have observed that there is no granger-causality from the oil price to the wheat price in the time intervals I investigate in which extends beyond January 2021. I am attributing the absence of granger-causality for the time intervals which extend beyond January 2021 to a negative demand shock for oil, likely caused by covid-19, as stringent travel restrictions reduced demand for fuels.

It is imperative that policymakers conduct though analyses before implementing new policies to avoid unintended secondary effects. I illustrated with an example in chapter three, a policy like the Renewable Fuel Standard (implemented in 2005 and strengthened in 2007 in the United States) is likely to increase the demand for corn and soybean for biofuel production, to meet the increased demand farmers substitute away the production of wheat (for food) in favor of corn and soybean (for fuel) (United States Environmental Protection Agency, 2023). The initial effects are reduced emissions which is likely the intended effect. However, the secondary effects of farmers substituting away from food towards fuel production, keep in mind that arable land can be considered as scarce, might lead to an inward shift of the supply curve of wheat as a secondary effect. An inward shift of supply will, in turn, lead to a reduced equilibrium quantity and an increased equilibrium price.

The potentially clashing United Nation sustainability goals, which are “zero hunger” (United Nations Sustainable Development Goals, 2023) and “clean and accessible energy for all” (United Nations Sustainable Development Goals, 2023) could potentially clash if biofuels are considered a part of the solution for their goal on energy, where the substitution mechanism becomes highly relevant. It could potentially put policymakers in a difficult situation where reducing emissions could have adverse effects on poorer people ability to afford food (this might not be true for people who are considered poor but which farms, they might enjoy the increased prices of their farm output, maybe worst for poor people living in urban areas). Said government policy can also adversely affect the goal of ending hunger through the input factor mechanism, if a policy increases the cost of petroleum-based fuel, this could increase the costs of food production, shifting the supply curve inwards, which in turn could increase the cost of agricultural commodities.

It is important for policymakers, governments, aid organizations, investors, and other stakeholders to be aware of the mechanisms described in this thesis as it could help them prepare and adjust to a new situation in advance to help mitigate the worst effects.

6.2. Limitations of the study

The assumptions I make throughout the thesis are limitations. If these assumptions are invalid, it can result in incorrect interpretation. In chapter three, I illustrate this with the example of a technological innovation in oil production known as fracking. The assumption I make is that the reduced production cost will cause the supply curve to shift outwards, resulting in an increase in the equilibrium quantity and a decrease in the equilibrium price. However, if the oil producers possess significant market power, they can counteract this mechanism.

The decisions I made during the process of collecting and selecting data for this thesis can be viewed as constraints, as they involve choices such as creating and using a price index instead of using nominal or real prices. Additionally, determining whether the baseline of the price index should be set at 100 in January 1990 or at a different point in time, considered an alternative frequency for data collection beyond one observation per month as well as the decision to start collecting data in

January 1990 may limit the study. Furthermore, deciding to accept FREDs explanations that both time series of commodity prices can be generalizable as global prices could pose a limitation if proven to be false.

Another limitation could be the fact that I have only investigated the relationship between the oil and wheat price, I do not have other variables in which could help explain the results.

Other limitations could be due to an unforeseen misspecification of the model I utilize, or incorrect interpretation of the BIC results. When conducting my analysis with lag lengths of 3, 4, 5, 6, and 8, I need to be cautious in interpreting the results. If the BIC suggests that either the lowest or highest number of lags is the optimal choice, I should pay special attention as it may lead to an incorrect conclusion. For instance, if the BIC indicates that 8 lags is the optimal choice it does not necessarily mean that 8 lags are suitable. Any number of lags equal to 8 or higher could potentially be the optimal choice, but as I maximum test with 8 lags I am unaware what I do not test for.

6.3. Recommendations and suggestions for further research

Firstly, I would recommend researching if oil price forecasts could predict either wheat prices or wheat price forecasts. This recommendation is based on my results where I have proven that the oil price can help predict the wheat price at the 5% significance level using the granger-causality test in several instances in this thesis.

Secondly, as discussed in the summary above, a likely reason which can explain why I do not prove granger-causality in time series which extend beyond January 2021 is due to a demand shock, which is in line with the results in the recent study of Umar et.al., (2021). I would highly recommend continuing the research on how supply versus demand shock affects the granger-causality as more data becomes available, to verify or refute our findings.

Thirdly, I would recommend furthering the research with the nonlinear specification model, which is based on Hamilton's framework (Hamilton (2003) and Hamilton (2009)) discussed in chapter four, ideally with other numbers of lags, compared to another more sophisticated model and for different oil and wheat indexes to verify or refute my findings.

Lastly, I suggest delving deeper into the mechanisms to enhance the body of literature on this subject matter.

7. References

Anderson, S. T., Kellogg, R., & Salant, S. W. (2018). Hotelling under Pressure. *Journal of Public Economics*, 984-1026.

André De Munter. (2023, April). *Factt Sheets on the European Union*. Retrieved from The European Parliament: https://www.europarl.europa.eu/factsheets/en/sheet/167/the-enlargement-of-the-union

Baumeister, C., & Kilian, L. (2014). Do oil price increases cause higher food prices? *Economic Policy*, 691-747.

Chen, P.-Y., Chang, C.-L., Chen, C.-C., & McAleer, M. (2012). Modelling the Effects of Oil Prices on Global Fertilizer Prices and Volatility. *Journal of Risk and Financial Management*, 78-114. doi:https://doi.org/10.3390/jrfm5010078

Chen, S.-T., Kuo, H.-I., & Chen, C.-C. (2010). Modeling the relationship between the oil price and global food prices. *Elsevier, Applied Energy*, 2517-2525.

Chen, S.-T., Kuo, H.-I., & Chen, C.-C. (2010). Modeling the relationship between the oil price and global food prices. *Applied Energy*, 2517-2525. doi:https://doi.org/10.1016/j.apenergy.2010.02.020

CME Group. (2019, August 27). *Wheat Futures and Options*. Retrieved from CME Group: https://www.cmegroup.com/education/brochures-and-handbooks/wheat-futures-and-options.html

Edelstein, P., & Kilian, L. (2007). Retail Energy Prices and Consumer Expenditures. *International Macroeconomics*, 1-48.

Enghiad, A., Danielle Ufer, A. M., & Thilmany, D. D. (2017). An Overview of Global Wheat Market Fundamentals in an Era of Climate Conserns . *International Journal of Agronomy*, 1-17.

Greenville, J. (2020). *Analysis of government support for Australian agricultural producers.* Canberra: Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES) research report.

Hamilton, J. D. (2003, April). What is an oil shock? *Journal of Econometrics*, 363-398. doi:https://doi.org/10.1016/S0304-4076(02)00207-5

Hamilton, J. D. (2009, May). Causes and Consequences of the Oil Shock of 2007-08. *National Bureau of Economic Research - NBER Working Paper Series*, 1-69. doi:10.3386/w15002

Hansen, R. G. (2013). *Mikroøkonomi - en kort innføring.* Oslo/Tuvalu: Gyldendal Norsk Forlag AS.

International Monetary Fund. (2019). *IMF Primary Commodity Prices.* Washington, D.C.: International Monetary Fund. Retrieved from file:///C:/Users/atlet/Downloads/pcps-technical-documentation%20(1).pdf

International Monetary Fund. (2023, June 21). *Global price of Wheat [PWHEAMTUSDM]*. Retrieved from FRED, Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/series/PWHEAMTUSDM

Kubursi, A. (2021, April 21). *World Economic Forum*. Retrieved from Oil crash explained: How are negative oil prices even possible?: https://www.weforum.org/agenda/2020/04/negative-oil-prices-covid19/

Lu, Y., Yang, L., & Liu, L. (2019). Volatility Spillovers between Crude Oil and Agricultural Commodity Markets since the Financial Crisis. *Sustainability, MDPI*, 1-12. doi:http://dx.doi.org/10.3390/su11020396

Mensi, W., Hammoudeh, S., Nguyen, D. K., & Yoon, S.-M. (2014). Dynamic spillovers among major energy and cereal commodity prices. *Energy Economics*, 225-243. doi:https://doi.org/10.1016/j.eneco.2014.03.004

National Bureau of Economic Research. (2022, August 15). *Business Cycle Dating Procedure: Frequently Asked Questions*. Retrieved from National Bureau of Economic Research: https://www.nber.org/research/business-cycle-dating/business-cycle-dating-procedure-frequently-asked-questions

National Bureau of Economic Research. (2023, March 14). *US Business Cycle Expansions and Contractions*. Retrieved from National Bureau of Economic Research: https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions

Nazlioglu, S. (2011). World oil and agricultural commodity prices: Evidence from nonlinear causality. *Energy Policy*, 2935-2943. doi:https://doi.org/10.1016/j.enpol.2011.03.001

Nazlioglu, S., & Soytas, U. (2012). Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis. *Energy Economics*, 1098-1104. doi:http://dx.doi.org/10.1016/j.eneco.2011.09.008

OECD. (2020). *Agricultural Poolicy Monitoring and Evaluation 2020.* Paris: OECD Publishing. doi:https://doi.org/10.1787/928181a8-en

Paris, A. (2018). On the link between oil and agricultural commodity prices: Do biofuels matter? *International Economics*, 48-60. doi:https://doi.org/10.1016/j.inteco.2017.12.003

Piringer, G., & Steinberg, L. J. (2008). Reevaluation of Energy Use in Wheat Production in the United States. *Journal of Industrial Ecology*, 149-167. doi:https://doi.org/10.1162/108819806775545420

Rafiq, S., & Bloch, H. (2016). Explaining Commodity Prices through Asymmetric Oil Shocks: Evidence from Nonlinear Models. *Resources Policy*, 1-42. doi:https://doi.org/10.1016/j.inteco.2017.12.003

Reboredo, J. C. (2012). Do food and oil prices co-move? *Energy Policy*, 456-467. doi:https://doi.org/10.1016/j.enpol.2012.06.035

Safa, M., Samarasinghe, S., & Mohssen, M. (2010). Determination of fuel consumption and indirect factors affecting it in wheat production in Canterbury, New Zealand. *Energy - Elsevier*, 5400-5405. doi:https://doi.org/10.1016/j.energy.2010.07.015

Saghaian, S. (2010). The Impact of the Oil Sector on Commodity Prices: Correlation or Causation? *Journal of Agricultural and Applied Economics*, 477-485.

Saghaian, S. H. (2010). The Impact of the Oil Sector on Commodity Prices: Correlation or Causation? *Journal of Agricultural and Applied Economics*, 477-485.

Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M., & Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Springer Link*, 291-317. doi:10.1007/s12571-013-0263-y

Statology (F-test). (2019, March 20). *Statology. Statistics. Simplified*. Retrieved from How to Read and Interpret a Regression Table: https://www.statology.org/read-interpret-regression-table/#:~:text=The%20f%20statistic%20is%20calculated%20as%20regression%20MS,the%20regression%20model%20as%20a%20whole%20is%20useful.

Statology. (2021, January 20). *Statology. Statistics. Simplified*. Retrieved from How to Detrend Data (With Examples): https://www.statology.org/detrend-data/

Statology, (Breusch-Godfrey Test). (2021, April 21). *Statology. Statistics. Simplified*. Retrieved from How to Perform a Breusch-Godfrey Test in Python: https://www.statology.org/breusch-godfrey-test-python/

Statology, (Breusch-Pagan Test). (2020, July 20). *Statology. Statistics. Simplified.* Retrieved from How to Perform a Breusch-Pagan Test in Python: https://www.statology.org/breusch-pagan-test-python/

Statology, (Dickey-Fuller Test). (2021, May 25). *Statology. Statistics. Simplified.* Retrieved from Augmented Dickey-Fuller Test in Python (With Examples): https://www.statology.org/dickey-fuller-test-python/

Stock, J. H., & Watson, J. W. (2015). *Introduction to Econometrics, Chapter 14 Introduction to Time Series Regression and Foreasting* (Update, 3rd Edition ed.). Harlow, England: Pearson Education Limited.

Stocker, M., Baffes, J., & Vorisek, D. (2018, January 18). *World Bank Blogs*. Retrieved from what triggered the oil price plunge of 2014-2016 and why it failed to deliver an economic impetus in eight charts: https://blogs.worldbank.org/developmenttalk/what-triggered-oil-price-plunge-2014-2016-and-why-it-failed-deliver-economic-impetus-eight-charts

The International Monetary Fund. (2023, July 10). *International Monetary Fund*. Retrieved from IMF Data, Access to Macroeconomic & Financial Data, Primary Commodity Price System (PCPS): https://data.imf.org/?sk=471dddf8-d8a7-499a-81ba-5b332c01f8b9&sid=1547558078595

U.S. Bureau of Labor Statistics. (2023, June 21). *Producer Price Index by Commodity: Fuels and Related Products and Power: Crude Petroleum (Domestic Production) [WPU0561]*. Retrieved from FRED, Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/series/WPU0561

Umar, Z., Gubareva, M., Naeem, M., & Akhter, A. (2021). Return and volatility transmission between oil price shocks and agricultural commodities. *PLOS ONE*, 1-18. doi:https://doi.org/10.1371/journal.pone.0246886

Umar, Z., Gubareva, M., Naeem, M., & Akhter, A. (2021, February 19). Return and Volatility Transmission Between Oil Price Shocks and Agricultural Commodities. *PLoS ONE (Public Library of Science), 16*(2), 1-18. doi:https://doi.org/10.1371/journal.pone.0246886

United Nations Sustainable Development Goals. (2023, September 20). *Ensure access to affordable, reliable, sustainable and modern energy*. Retrieved from United Nations: https://www.un.org/sustainabledevelopment/energy/

United Nations Sustainable Development Goals. (2023, September 20). *Goal 2: Zero Hunger*. Retrieved from United Nations: https://www.un.org/sustainabledevelopment/hunger/

United States Department of Agriculture - Economic Research Service. (2023, January 5). *Commodity Costs and Returns*. Retrieved from United States Department of Agriculture - Economic Research Service: https://www.ers.usda.gov/data-products/commodity-costs-and-returns/commodity-costs-and-returns/#Historical%20Costs%20and%20Returns:%20Wheat

United States Department of Agriculture - Foreign Agricultural Service. (2023, July 29). *PS&D - Production, Supply and Distribution*. Retrieved from USDA - United States Department of Agriculture - Foreign Agricultural Service: https://apps.fas.usda.gov/psdonline/app/index.html#/app/advQuery

United States Department of Agriculture. (2014). *U.S. Standards.* Agricultural Marketing Service - Federal Grain Inspection Service. Washington: USDA - AMS - FGIS. Retrieved from https://www.ams.usda.gov/sites/default/files/media/WheatStandards.pdf

United States Environmental Protection Agency. (2023, February 10). *United States Environmental Protection Agency, EPA*. Retrieved from Overview for Renewable Fuel Standard: https://www.epa.gov/renewable-fuel-standard-program/overview-renewable-fuel-standard

Wooldridge, J. M. (2021). *Introductory Econometrics - A Modern Approach, chapter 11 Further Issues in Using OLS with Time Series Data* (7th Edition ed.). Ashford, United Kingdom: Cengage Learning.

Zmami, M., & Ben-Salha, O. (2019). Does Oil Price Drive World Food Prices? Evidence from Linear and Nonlinear ARDL Modeling. *Economies*, 1-18. doi:http://dx.doi.org/10.3390/economies7010012

Appendix 1

The data which is gathered, compiled, and presented in chapter 2.4.2. Global Wheat Exports are retrieved from the United States Department of Agriculture, USDA, are described in detail in this appendix. As global wheat data, to my knowledge, is not available in a time series format, I have manually downloaded, compiled, and analyzed annual wheat export data from the USDAs website to create an overview and visualize trends and export volumes from 1990 to 2022.

To collect the necessary data, I have checked for *Wheat* as the commodity, *Exports* as the attributes, *World Total* as countries and each year in the time span from *1990 to 2022* for the market years in question in the query. I thereafter chose the top 10 exporting countries per year and downloaded the results from each query (year). I compiled the results from my queries into one time series to analyze trends and aggregated exports by country or region*. The region I refer to is Europe, as it starts out as *EU-15* and evolves into *The European Union* in 1999. According to a Fact Sheet on the European Union published by the European Parliament, 16 European countries have joined the European Union in the time period we're investigating, including medium sized wheat exporting countries as Hungary (among top ten exporting countries in several executive years in the 1990s) and Czechia (also known as the Czech Republic) of which both countries officially joined The European Union in 2004, Romania which officially joined The European Union in 2007 as well as the United Kingdom which officially left The European Union in 2020 (André De Munter, 2023). Please note that the export data from the European Union is a snapshot of total exports each year and it is not adjusted backwards or forward for countries before they join the European Union (especially note Hungary, the Czech Republic and Romania before joining the European Union before 2004 and 2007, respectively) or the loss of export from the United Kingdom as they exited the European Union (officially in 2020). With this information in mind, the wheat export from Europe is likely underestimated, especially from 1990 to 2004.

Upon compiling and analyzing the ten largest wheat exporting countries or regions in the world I made the choice to limit my focus to the top five exporting countries per year and to gather any missing data from USDA. In the period I investigated, 1990 to 2022, I discovered that the United States, the European Union, Canada, Australia, Argentina, Russia, Ukraine, Turkey, Kazakhstan, and Others made it to the top five exporting countries in at least one year. As Turkey and Kazakhstan only made it to the top five exporting countries once each and Others was unspecified, those three got eliminated from the list. Upon comparing the seven largest wheat exporting countries to the total world exports, the seven largest exporting countries accounted for roughly 85% of global wheat export from 1990 through 2022, underpinning my decision to limit my analysis to the largest exporting countries.

Master Thesis - Atle Torgersen - Fall 2023

Can the Oil Price predict the Wheat Price?

Following is a copy of the code I used for one example, namely for time series A: 1990-01-01:2020-01-01 with eight lags.

In [ ]:
```
# Import packages:

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
import statsmodels.formula.api as smf
import statsmodels.stats.api as sms
import statsmodels.stats.diagnostic as smd
import scipy.stats as stats
import linearmodels.iv as iv
import seaborn as sns
import sys
import os
from itertools import combinations
import datetime as dt
import pandas_datareader as pdr


import statsmodels as sm
import scipy.stats as stats
import wooldridge as woo
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.stattools import adfuller
from statsmodels.stats.stattools import durbin_watson
from statsmodels.tsa.stattools import grangercausalitytests
from statsmodels.tsa.api import VAR, VARMAX
from statsmodels.tsa.vector_ar.vecm import coint_johansen

%matplotlib inline
```

In [ ]:
```
# Problem 1

# Prepare the dataframe

# Setting the file path for getting the data
path = "/Users/atlet/.spyder-py3/M30-ECON- 23V Master Thesis"
os.chdir(path)

# Upload the Oil data (WPU0561)
Oildata = pd.read_csv('WPU0561 Indexed 1990.csv',parse_dates = [0])
Oildata['DATE'] = Oildata['DATE'].dt.to_period('M')
Oildata = Oildata.set_index('DATE').sort_index()
#Oildata = Oildata.resample('M').last()

# Upload the Wheat data (PWHEAMTUSDM)
Wheatdata = pd.read_csv('PWHEAMTUSDM Indexed 1990.csv',parse_dates = [0])
Wheatdata['DATE'] = Wheatdata['DATE'].dt.to_period('M')
Wheatdata = Wheatdata.set_index('DATE').sort_index()
#Wheatdata = Wheatdata.resample('M').last()


# Merge Data
```

```
data = pd.merge(Oildata, Wheatdata, how = 'outer', on = 'DATE')

# Rename columns
data = data.rename(columns={"WPU0561 Indexed 1990":"Oil"})
data = data.rename(columns={"PWHEAMTUSDM Indexed 1990":"Wheat"})

# Get new variables

# Hamilton’s oil price and oil price change variables
data['jOil'] = 100*np.log(data['Oil'])
data['djOil'] = data['jOil'] - data['jOil'].shift(1)
data['djOil'] = data['djOil'].fillna(0)

# cumulative change in oil price
data['cdop'] = data['djOil'].cumsum()

# maximum oil price change last three years
data['maxp'] = data['cdop'].rolling(window=12,min_periods=1).max()
data['maxp'] = data['maxp'].fillna(0)

# Hamilton’s oil price change measure
data['jpmax'] = data['cdop'] - data['maxp'].shift(1)
data.loc[data['jpmax'] < 0,'jpmax'] = 0

# New Wheat variables
data['jWheat'] = 100*np.log(data['Wheat'])
data['djWheat'] = data['jWheat'].diff(1)

# Write new datafile to csv
data.to_csv('hamilton_OilWheatdata9020.csv')
```

In [ ]:

```
# 2.1. Summarize the Wheat and oil price data

# Read the data from the csv file we just made:
data = pd.read_csv(r"C:\Users\atlet\.spyder-py3\M30-ECON- 23V Master Thesis\hamilton_
                   index_col=[0])

# Subset the data:
data_9020 = data["1990-01-01":"2023-04-01"]
data_9020
```

In [ ]:

```
# 2.2. Create one graph showing the Oil and Wheat price time series (1990-01-01:2023

# ADDED 21.06.2023

# Plot Oil and Wheat price:

fig, ax = plt.subplots(2,1,figsize=(12,8))
fig.suptitle('Oil and Wheat Prices 1990-01-01:2020-01-01 (Index = 100 in January 199

data_9020['Oil'].plot(ax=ax[0],c='r')
ax[0].set_ylabel('Oil price (index = 100 in January 1990)')
ax[0].grid()
ax[0].set_xlabel('Year')

data_9020['Wheat'].plot(ax=ax[1],c='b')
ax[1].set_ylabel('Wheat Price (index = 100 in January 1990')
ax[1].grid()
ax[1].set_xlabel('Year')

fig.tight_layout()
fig.savefig('./Oil and Wheat9020.png')
```

In [ ]:

```
# 2.3. Create one graph showing (the log version of) theses two time series (1990-01

# Plot (log) Wheat- and Oil price:

fig, ax = plt.subplots(2,1,figsize=(12,8))
fig.suptitle('Oil and Wheat Prices 1990 (log)', fontsize=18)

data_9020['jOil'].plot(ax=ax[0],c='r')
ax[0].set_ylabel('(log) oil price')
ax[0].grid()
ax[0].set_xlabel('Year')

data_9020['jWheat'].plot(ax=ax[1],c='b')
ax[1].set_ylabel('(log) Wheat Price')
ax[1].grid()
ax[1].set_xlabel('Year')

fig.tight_layout()
fig.savefig('./log of Wheat and oil9020.png')
```

In [ ]:

```
# 2.4. Create another graph showing the changes in these two time series for the tim

# Plot (log) Wheat- and Oil price GROWTH:

fig, ax = plt.subplots(2,1,figsize=(12,8))
fig.suptitle('Growth in Oil and Wheat Prices from 1990 (log)', fontsize=18)

data_9020['djOil'].plot(ax=ax[0],c='r')
ax[0].set_ylabel('% Change in (log) oil price')
ax[0].grid()
ax[0].set_xlabel('Year')

data_9020['djWheat'].plot(ax=ax[1],c='b')
ax[1].set_ylabel('% Change in (log) wheat Prices')
ax[1].grid()
ax[1].set_xlabel('Year')

fig.tight_layout()
fig.savefig('./Change in (log) of wheat and oil9020.png')
```

In [ ]:

```
# Problem 3.1.

# create lagged variables from the oil- and wheat price change, which is jpmax and d

# Create djpmax:
data["djpmax"] = data["jpmax"] - data["jpmax"].shift(1)

# Create lagged variables for AR(4) on jpmax:
data['jpmax_L1'] = data['jpmax'].shift(1)
data['jpmax_L2'] = data['jpmax'].shift(2)
data['jpmax_L3'] = data['jpmax'].shift(3)
data['jpmax_L4'] = data['jpmax'].shift(4)
data['jpmax_L5'] = data['jpmax'].shift(5)
data['jpmax_L6'] = data['jpmax'].shift(6)
data['jpmax_L7'] = data['jpmax'].shift(7)
data['jpmax_L8'] = data['jpmax'].shift(8)

# Create lagged variables for AR(4) on djWheat:
data['djWheat_L1'] = data['djWheat'].shift(1)
```

```
data['djWheat_L2'] = data['djWheat'].shift(2)
data['djWheat_L3'] = data['djWheat'].shift(3)
data['djWheat_L4'] = data['djWheat'].shift(4)
data['djWheat_L5'] = data['djWheat'].shift(5)
data['djWheat_L6'] = data['djWheat'].shift(6)
data['djWheat_L7'] = data['djWheat'].shift(7)
data['djWheat_L8'] = data['djWheat'].shift(8)
```

In [ ]:

```
# Problem 3.2.1.

# Test for stationarity of Wheat:

print("Observations of Dickey-fuller test on GDP")
dftest_1 = adfuller(data['Wheat'],autolag='AIC')
dfoutput_1=pd.Series(dftest_1[0:4],index=['Test Statistic','p-value','#lags used','n
for key,value in dftest_1[4].items():
    dfoutput_1['critical value (%s)'%key]= value
print(dfoutput_1)
print()

if dftest_1[1] <= 0.05:
    print("Strong evidence against the null hypotehsis")
    print("Reject the null hypothesis")
    print("Data has no unit root and is stationary")
else:
    print("No evidence against the null hypothesis")
    print("Fail to reject the null hypothesis")
    print("Data has a unit root and is non-stationary")
```

In [ ]:

```
# Problem 3.2.2.
# Test for stationarity of Oil:

print("Observations of Dickey-fuller test on Oil price")
dftest_2 = adfuller(data['Oil'],autolag='AIC')
dfoutput_2=pd.Series(dftest_2[0:4],index=['Test Statistic','p-value','#lags used','n
for key,value in dftest_2[4].items():
    dfoutput_2['critical value (%s)'%key]= value
print(dfoutput_2)
print()

if dftest_2[1] <= 0.05:
    print("Strong evidence against the null hypotehsis")
    print("Reject the null hypothesis")
    print("Data has no unit root and is stationary")
else:
    print("No evidence against the null hypothesis")
    print("Fail to reject the null hypothesis")
    print("Data has a unit root and is non-stationary")
```

In [ ]:

```
# Problem 3.2.3.
# Test for stationarity of jpmax

print("Observations of Dickey-fuller test on Oil price change (jpmax)")
dftest_3 = adfuller(data['jpmax'].dropna(),autolag='AIC')
dfoutput_3=pd.Series(dftest_3[0:4],index=['Test Statistic','p-value','#lags used','n
for key,value in dftest_3[4].items():
    dfoutput_3['critical value (%s)'%key]= value
print(dfoutput_3)
print()
```

```python
if dftest_3[1] <= 0.05:
        print("Strong evidence against the null hypothesis")
        print("Reject the null hypothesis")
        print("Data has no unit root and is stationary")
else:
        print("No evidence against the null hypothesis")
        print("Fail to reject the null hypothesis")
        print("Data has a unit root and is non-stationary")
```

In [ ]:

```python
# Problem 3.2.4. ADDED!
# Test for stationarity of djwheat

print("Observations of Dickey-fuller test on Wheat price change (djwheat)")
dftest_3 = adfuller(data['djWheat'].dropna(),autolag='AIC')
dfoutput_3=pd.Series(dftest_3[0:4],index=['Test Statistic','p-value','#lags used','n
for key,value in dftest_3[4].items():
    dfoutput_3['critical value (%s)'%key]= value
print(dfoutput_3)
print()

if dftest_3[1] <= 0.05:
        print("Strong evidence against the null hypothesis")
        print("Reject the null hypothesis")
        print("Data has no unit root and is stationary")
else:
        print("No evidence against the null hypothesis")
        print("Fail to reject the null hypothesis")
        print("Data has a unit root and is non-stationary")
```

In [ ]:

```python
# Problem 4.1.

# Summarize the wheat and oil price data (including the lagged variables)

data_9020_8 = data["1990-01-01":"2020-01-01"].copy()
data_9020_8
```

In [ ]:

```python
# Problem 4.2.

# AR(8) model
ar8mod = smf.ols('djWheat ~ djWheat_L1 + djWheat_L2 + djWheat_L3 + djWheat_L4 + djWh
                jpmax_L1 + jpmax_L2 + jpmax_L3 + jpmax_L4 + jpmax_L5 + jpmax_L6 + jp
                 ,data=data_9020_8)
ar8res = ar8mod.fit(cov_type="HC3")

print(ar8res.summary())
```

In [ ]:

```python
# Problem 4.2.1.

# summary of in-sample performance

print()
print("Model     AIC      BIC")
print("-----------------------")
print("ar(8) %8.2f %8.2f" % (ar8res.aic, ar8res.bic))
print()

jWheat = data["jWheat"].values
```

In [ ]:

```python
# Problem 4.3.

# Test the residuals for serial correlation:

# Predicted values and residuals
data_9020_8['ar8_fitted'] = ar8res.fittedvalues
data_9020_8['ar8_res'] = ar8res.resid
data_9020_8['ar8_res_lag'] = data_9020_8['ar8_res'].shift(1)

# Built-in Breusch-Godfrey test
# (PS: alternatively use a Breusch-Pagan test made for AR(4) models):
bgx,bgxpv,bgf,bgfpv = sm.stats.diagnostic.acorr_breusch_godfrey(ar8res,1)
print("B-G Test for Serial Correlation")
print(bgx)
print(bgxpv)
print(bgf)
print(bgfpv)

bgmod1 = smf.ols(formula=('ar8_res ~ djWheat_L1 + djWheat_L2 + djWheat_L3 + djWheat_
                          + djWheat_L7 + djWheat_L8 + jpmax_L1 + jpmax_L2 + jpmax_L3
                          + jpmax_L7 + jpmax_L8 + ar8_res_lag'),data=data_9020_8)
bgr1 = bgmod1.fit(cov_type="HC3")
print(bgr1.summary())

bgmod2 = smf.ols(formula=('ar8_res ~ ar8_res_lag'),data=data_9020_8)
bgr2 = bgmod2.fit(cov_type="HC3")
print(bgr2.summary())
```

In [ ]:

```python
# Problem 4.3.

# Test for Heteroskedasticity:

# Using Breusch-Pagan test:
data_9020_8["u_sqr"] = ar8res.resid ** 2
bphmod = smf.ols(formula="u_sqr ~ djWheat + djWheat_L1 + djWheat_L2 + djWheat_L3 + d
                                + djWheat_L7 + djWheat_L8 + jpmax_L1 + jpmax_L2 + jp
                                + jpmax_L6 + jpmax_L7 + jpmax_L8", data=data_9020_8)
bphres = bphmod.fit(cov_type="HC1")
bph_lm = bphres.nobs * bphres.rsquared
bph_pv = 1 - stats.chi2.cdf(x=bph_lm,df= bphmod.df_model)
print()
print("Testing for heteroskedasticity")
print("Aux regression, R-squared = {}" .format(bphres.rsquared))
print("Aux regression, d of free = {}" .format(bphmod.df_model))
print("Number of observations    = {}" .format(bphres.nobs))

print()
print("Breusch-Pagan LM-test: {}, p-value={}" .format(round(bph_lm,2), round(bph_pv,
print()
```

In [ ]:

```python
# Problem 4.4.

# Test for Granger causality of oil price change:

hoil = ["jpmax_L1 = 0", "jpmax_L2 = 0", "jpmax_L3 = 0", "jpmax_L4 = 0", "jpmax_L5 =
        "jpmax_L8 = 0"]
fhoil = ar8res.f_test(hoil)

print("Testing if oil price change Granger causes GDP change")
```

```
print("F-stat : {}".format(fhoil.statistic[0][0]))
print("p-value: {}".format(fhoil.pvalue))
```

In [ ]:
```
# Problem 5.2.

# Estimate a pure AR(4) model:
ar8wheatmod = smf.ols('djWheat ~ djWheat_L1 + djWheat_L2 + djWheat_L3 + djWheat_L4 +
                        + djWheat_L8',data=data_9020_8)
ar8wheatres = ar8wheatmod.fit(cov_type="HC3")

print(ar8wheatres.summary())
```

In [ ]:
```
# Problem 5.2.1.

# Summary of in-sample performance

print()
print("Model      AIC       BIC")
print("-----------------------")
print("ar(8) %8.2f %8.2f" % (ar8wheatres.aic, ar8wheatres.bic))
print()

jgdp = data["jWheat"].values
```

In [ ]:
```
# Problem 5.3.

# Test the residuals for serial correlation:

# Predicted values and residuals
data_9020_8['ar8wheat_fitted'] = ar8wheatres.fittedvalues
data_9020_8['ar8wheat_res'] = ar8wheatres.resid
data_9020_8['ar8wheat_res_lag'] = data_9020_8['ar8wheat_res'].shift(1)

# Built-in Breusch-Godfrey test
# (PS: alternatively use a Breusch-Pagan test made for AR(4) models):
bgx,bgxpv,bgf,bgfpv = sm.stats.diagnostic.acorr_breusch_godfrey(ar8wheatres,1)
print("B-G Test for Serial Correlation")
print(bgx)
print(bgxpv)
print(bgf)
print(bgfpv)
```

In [ ]:
```
# Problem 5.3.

# Test for Heteroskedasticity:

data_9020_8["ugdp_sqr"] = ar8wheatres.resid ** 2
bphmod2 = smf.ols(formula="ugdp_sqr ~ djWheat + djWheat_L1 + djWheat_L2 + djWheat_L3
                    + djWheat_L6 + djWheat_L7 + djWheat_L8", data=data_9020_8)
bphres2 = bphmod2.fit(cov_type="HC1")
bph_lm2 = bphres2.nobs * bphres2.rsquared
bph_pv2 = 1 - stats.chi2.cdf(x=bph_lm2,df= bphmod2.df_model)
print()
print("Testing for heteroskedasticity")
print("Aux regression, R-squared = {}" .format(bphres2.rsquared))
print("Aux regression, d of free = {}" .format(bphmod2.df_model))
print("Number of observations    = {}" .format(bphres2.nobs))

print()
```

```
print("Breusch-Pagan LM-test: {}, p-value={}" .format(round(bph_lm2,2), round(bph_pv
print()
```

In [ ]:

```
# Problem 5.4.

# Test for Granger-Causality on the dependant variables own lags (wheat price change

hwheat = ["djWheat_L1 = 0", "djWheat_L2 = 0", "djWheat_L3 = 0", "djWheat_L4 = 0"]
fhwheat = ar4res.f_test(hwheat)

print("Testing if oil price change Granger causes GDP change")
print("F-stat : {}".format(fhwheat.statistic[0][0]))
print("p-value: {}".format(fhwheat.pvalue))
```

In [ ]:

```
# Problem 6

# Set ID index tobe able to convert DATE to quarterly observations

#data = data.reset_index()

# Set time to quarterly observations and set DATE as index

#data['DATE'] = data['DATE'].dt.to_period('M')
#data = data.set_index('DATE').sort_index()

# Subsetting to create one estimation and one validation dataset:

# Estimation period, use Hamilton data, subset for period 1949Q2-2007Q4
data1990_2010est = data[(data.index >= pd.Period('1990-01')) & (data.index <= pd.Per

# Validation period, use Hamilton data, subset for period 2007Q4-2010Q4
data2010_2020val = data[(data.index >= pd.Period('2010-01')) & (data.index <= pd.Per
```

In [ ]:

```
# Problem 6

# Hamilton data - subset 1949Q2 - 2007Q3:
data_2007Q3 = data[(data.index >= '1990.01.01') & (data.index <= '2009.12.01')].copy
#print(data_2007Q3)

# Creating a subset of our data - 1949Q2 - 2007Q3
data_2007Q3 = data["1990-01-01":"2010-01-01"].copy()
data_2007Q3
```

In [ ]:

```
# Problem 6

# Creating the new AR(4) model
ar4gdpmod2 = smf.ols('djWheat ~ djWheat_L1 + djWheat_L2 + djWheat_L3 + djWheat_L4',d
ar4gdpres2 = ar4gdpmod2.fit(cov_type="HC1")

print(ar4gdpres2.summary())
```

In [ ]:

```
# Problem 6

# Summary of in-sample performance

print()
print("Model     AIC       BIC")
print("-----------------------")
```

```
print("ar(4) %8.2f %8.2f" % (ar4gdpres2.aic, ar4gdpres2.bic))
print()

jgdp = data["jWheat"].values
```

In [ ]:

```
jWheat = data["jWheat"].values
```

In [ ]:

```
# Problem 6

# Forecasting with AR(4) model

# Keep the forecasts here
lghat4 = np.zeros(len(data))
dghat4 = np.zeros(len(data))

# Loop over observations
s = 0.0
i = 0
n = 0
for p, v in data.iterrows():
    if p < pd.Period("2011-01"):
        dghat4[i] = v["djWheat"]
        lghat4[i] = v["jWheat"]
    else:
        dghat4[i] = (ar4gdpres2.params[0] +
                     ar4gdpres2.params[1]*dghat4[i-1] +
                     ar4gdpres2.params[2]*dghat4[i-2] +
                     ar4gdpres2.params[3]*dghat4[i-3] +
                     ar4gdpres2.params[4]*dghat4[i-4])
        lghat4[i] = lghat4[i-1] + dghat4[i]

# Prediction error
    s += (jWheat[i] - lghat4[i])**2
    n += 1
    i +=1

# Keep the forecast
data["lghat4"] = lghat4

# Forecast error
rmse4 = np.sqrt(s/n)


# Plot graph of forecasts
dhx6 = data[(data.index >= pd.Period('20011-01')) & (data.index <= pd.Period('2020-0

fig, ax = plt.subplots(figsize=(9, 6))
fig.suptitle('Forecasting GDP for US 2007Q4 - 2010Q4', fontsize=18)
dhx6['jWheat'].plot(c='k', label='Actual')
dhx6['lghat4'].plot(c='g', label='Predicted from AR(4)')
ax.set_ylabel('(log) GDP')
ax.set_xlabel('Year')
ax.legend(loc='upper left')
ax.grid()
fig.tight_layout()
```

In [ ]:

```
# Problem 7
# Create a multiperiod forecast starting in 2007:Q4 and ending in 2010:Q4 for GDP (L
# the Hamilton model conditional upon the observed oil price change measure.
```

```
# Estimate Hamilton AR(4) model -> here you just insert the Hamilton variables in ad
hmod = smf.ols('djgdp ~ djgdp_L1 + djgdp_L2 + djgdp_L3 + djgdp_L4 + jpmax_L1 + jpmax
               data=data_2007Q3)
hres = hmod.fit(cov_type="HC3")
print(hres.summary())
```

In [ ]:

```
# Problem 7

# summary of in-sample performance
print()
print('Model      AIC      BIC')
print('-----------------------')
print('AR(4) %8.2f %8.2f' % (hres.aic, hres.bic))
print()
```

In [ ]:

```
# Problem 7

# Observed log GDP
jgdph = data['jgdp'].values
```

In [ ]:

```
# Problem 7

# Keep the forecasts here
jghat = np.zeros(len(data))
dghat = np.zeros(len(data))

# PS: ar4 from Hamilton AR(4) model in problem 4.1.

# Loop over observations
i = 0
for t, v in data.iterrows():
    if t < pd.Period('2007Q4'):
        jghat[i] = v['jgdp']
        dghat[i] = v['djgdp']
    else:
        dghat[i] = (hres.params[0] +
                    hres.params[1]*dghat[i-1] +
                    hres.params[2]*dghat[i-2] +
                    hres.params[3]*dghat[i-3] +
                    hres.params[4]*dghat[i-4] +
                    hres.params[5]*v['jpmax_L1'] +
                    hres.params[6]*v['jpmax_L2'] +
                    hres.params[7]*v['jpmax_L3'] +
                    hres.params[8]*v['jpmax_L4'])
        jghat[i] = jghat[i-1] + dghat[i]

    i += 1


# Keep the forecast
data['jghat'] = jghat

# Plot graph of forecasts
dhx7 = data[(data.index >= pd.Period('2011-01')) & (data.index <= pd.Period('2020-01

fig, ax = plt.subplots(figsize=(9, 6))
fig.suptitle('Forecasting GDP for US conditional on Oil Price 2007Q4 - 2010Q4', font
dhx7['jgdp'].plot(c='k', label='Actual')
dhx7['jghat'].plot(c='r', label='Predicted given oil prices')
ax.set_ylabel('(log) GDP')
```

```
ax.set_xlabel('Year')
ax.legend(loc='upper left')
ax.grid()
fig.tight_layout()
```

In [ ]:

```
# Problem 8
# Create a graph with actual GDP and your two forecasts for the period 2007:Q3-2008:(
# Hamilton’s graph. How well are you doing in replicating the results?


dfvalx8 = data[(data.index >= pd.Period('2007Q3')) & (data.index <= pd.Period('2008Q
hamvalx8 = data[(data.index >= pd.Period('2007Q3')) & (data.index <= pd.Period('2008

fig, ax = plt.subplots(figsize=(9,6))
fig.suptitle("Forecasting US GDP for 2007Q3 - 2008Q4", fontsize=18)
dfvalx8['jgdp'].plot(c="k",label="Actual GDP")
dfvalx8['lghat4'].plot(c="g",label="Predicted from AR(4)")
hamvalx8['jghat'].plot(c="r",label="Predicted given oil prices")
ax.set_ylabel("Log of Real GDP")
ax.set_xlabel("Date")
ax.legend(loc="upper left")
ax.grid()
fig.tight_layout()
```

In [ ]:

```
# Problem 9

dfvalx9 = data[(data.index >= pd.Period('2007Q3')) & (data.index <= pd.Period('2010Q
hamvalx9 = data[(data.index >= pd.Period('2007Q3')) & (data.index <= pd.Period('2010

fig, ax = plt.subplots(figsize=(9,6))
fig.suptitle("Forecasting US GDP for 2007Q3 - 2010Q4", fontsize=18)
dfvalx9["jgdp"].plot(c="k",label="Actual")
dfvalx9["lghat4"].plot(c="g",label="Predicted from AR(4)")
hamvalx9["jghat"].plot(c="r",label="Predicted given oil prices")
ax.set_ylabel("Log of Real GDP")
ax.set_xlabel("Date")
#ax.set_ylim([960,978])
ax.legend(loc="upper left")
ax.grid()
fig.tight_layout()
```

In [ ]:

```
# Problem 10.1.

# Adjust the data length
dh_updated = data[(data.index >= '1949.04.01') & (data.index <= '2016.07.01')].copy(

# Re-estimate AR(4) model
ar4_gdpmod_updated = smf.ols('djgdp ~ djgdp_L1 + djgdp_L2 + djgdp_L3 + djgdp_L4', da
ar4_gdpres_updated = ar4_gdpmod_updated.fit(cov_type='HC3')

# Re-estimate Hamilton model
var4ham_mod_updated = smf.ols('djgdp ~ djgdp_L1 + djgdp_L2 + djgdp_L3 + djgdp_L4 + j
var4ham_res_updated = var4ham_mod_updated.fit(cov_type='HC3')
```

In [ ]:

```
# Problem 10.2.

# Compare the Original Hamilton Model (data from 1949Q2 to 2001Q1) to the updated da
print("\nOriginal Hamilton Model:")
print(ar4res.summary())
```

```
print("\nUpdated Hamilton Model:")
print(var4ham_res_updated.summary())

# Compare the Original AR(4) Model (data from 1949Q2 to 2001Q1) to the updated data |
print("Original AR(4) Model:")
print(ar4gdpres.summary())
print("\nUpdated AR(4) Model:")
print(ar4_gdpres_updated.summary())
```

In [ ]:

```
# Problem 10.3.

# Modify the forecasting loops to start the forecast from 2017:Q1 for both models
# Reinitialize lghat4 and jghat arrays to keep updated forecasts
lghat4_updated_2017 = np.zeros(len(data))
dghat4_updated_2017 = np.zeros(len(data))
jghat_updated_2017 = np.zeros(len(data))
dghat_updated_2017 = np.zeros(len(data))

i = 0
for t, v in data.iterrows():
    if t < pd.Period('2017Q1'):
        lghat4_updated_2017[i] = v['jgdp']
        dghat4_updated_2017[i] = v['djgdp']
        jghat_updated_2017[i] = v['jgdp']
        dghat_updated_2017[i] = v['djgdp']
    else:
        dghat4_updated_2017[i] = (ar4_gdpres_updated.params[0] +
                                  ar4_gdpres_updated.params[1]*dghat4_updated_2017[i
                                  ar4_gdpres_updated.params[2]*dghat4_updated_2017[i
                                  ar4_gdpres_updated.params[3]*dghat4_updated_2017[i
                                  ar4_gdpres_updated.params[4]*dghat4_updated_2017[i
        lghat4_updated_2017[i] = lghat4_updated_2017[i-1] + dghat4_updated_2017[i]

        dghat_updated_2017[i] = (var4ham_res_updated.params[0] +
                                 var4ham_res_updated.params[1]*dghat_updated_2017[i-
                                 var4ham_res_updated.params[2]*dghat_updated_2017[i-
                                 var4ham_res_updated.params[3]*dghat_updated_2017[i-
                                 var4ham_res_updated.params[4]*dghat_updated_2017[i-
                                 var4ham_res_updated.params[5]*v['jpmax_L1'] +
                                 var4ham_res_updated.params[6]*v['jpmax_L2'] +
                                 var4ham_res_updated.params[7]*v['jpmax_L3'] +
                                 var4ham_res_updated.params[8]*v['jpmax_L4'])
        jghat_updated_2017[i] = jghat_updated_2017[i-1] + dghat_updated_2017[i]
    i += 1

# Add updated forecasts to the data
data['lghat4_updated_2017'] = lghat4_updated_2017
data['jghat_updated_2017'] = jghat_updated_2017
```

In [ ]:

```
# Problem 10.4.

# Create a new dataframe for the specified range
df_2016_2019 = data[(data.index >= pd.Period('2016Q4')) & (data.index <= pd.Period('

# Plot the graph
fig, ax = plt.subplots(figsize=(9, 6))
fig.suptitle('Forecasting GDP for US 2016Q4 - 2019Q4', fontsize=18)
df_2016_2019['jgdp'].plot(c='k', label='Actual')
df_2016_2019['lghat4'].plot(c='g', label='Predicted from AR(4)')
df_2016_2019['jghat'].plot(c='r', label='Predicted given oil prices')
ax.set_ylabel('(log) GDP')
ax.set_xlabel('Date')
```

```python
ax.legend(loc='upper left')
ax.grid()
fig.tight_layout()
```

In [ ]:

```python
# Problem 10.5.

# Modify the forecasting loops to start the forecast from 2019:Q4 for both models
# Reinitialize lghat4 and jghat arrays to keep updated forecasts
lghat4_updated = np.zeros(len(data))
dghat4_updated = np.zeros(len(data))
jghat_updated = np.zeros(len(data))
dghat_updated = np.zeros(len(data))

i = 0
for t, v in data.iterrows():
    if t < pd.Period('2019Q4'):
        lghat4_updated[i] = v['jgdp']
        dghat4_updated[i] = v['djgdp']
        jghat_updated[i] = v['jgdp']
        dghat_updated[i] = v['djgdp']
    else:
        dghat4_updated[i] = (ar4_gdpres_updated.params[0] +
                              ar4_gdpres_updated.params[1]*dghat4_updated[i-1] +
                              ar4_gdpres_updated.params[2]*dghat4_updated[i-2] +
                              ar4_gdpres_updated.params[3]*dghat4_updated[i-3] +
                              ar4_gdpres_updated.params[4]*dghat4_updated[i-4])
        lghat4_updated[i] = lghat4_updated[i-1] + dghat4_updated[i]

        dghat_updated[i] = (var4ham_res_updated.params[0] +
                            var4ham_res_updated.params[1]*dghat_updated[i-1] +
                            var4ham_res_updated.params[2]*dghat_updated[i-2] +
                            var4ham_res_updated.params[3]*dghat_updated[i-3] +
                            var4ham_res_updated.params[4]*dghat_updated[i-4] +
                            var4ham_res_updated.params[5]*v['jpmax_L1'] +
                            var4ham_res_updated.params[6]*v['jpmax_L2'] +
                            var4ham_res_updated.params[7]*v['jpmax_L3'] +
                            var4ham_res_updated.params[8]*v['jpmax_L4'])
        jghat_updated[i] = jghat_updated[i-1] + dghat_updated[i]
    i += 1

# Add updated forecasts to the data
data['lghat4_updated'] = lghat4_updated
data['jghat_updated'] = jghat_updated
```

In [ ]:

```python
# Problem 10.6.

# Create a new dataframe for the specified range
df_2018_2022 = data[(data.index >= pd.Period('2018Q4')) & (data.index <= pd.Period('

# Plot the graph
fig, ax = plt.subplots(figsize=(9, 6))
fig.suptitle('Forecasting GDP for US 2018Q4 - 2022Q4', fontsize=18)
df_2018_2022['jgdp'].plot(c='k', label='Actual')
df_2018_2022['lghat4_updated'].plot(c='g', label='Predicted from AR(4)')
df_2018_2022['jghat_updated'].plot(c='r', label='Predicted given oil prices')
ax.set_ylabel('(log) GDP')
ax.set_xlabel('Date')
ax.legend(loc='upper left')
ax.grid()
fig.tight_layout()
```

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway