**CHARLES UNIVERSITY**

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

**Bachelor thesis**

**2019** **Matěj Novotný**

**CHARLES UNIVERSITY**

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

**Matěj Novotný**

# Estimating the Relationship between Food, Fuel and Biofuel Prices

*Bachelor thesis*

Prague 2019

**Author**: Matěj Novotný

**Supervisor**: prof. Ing. Karel Janda, M.A., Dr., Ph.D.

**Academic Year**: 2018/2019

# Bibliographic note

Novotný, M. (2019), *Estimating the Relationship between Food, Fuel and Biofuel Prices.* Bachelor thesis (Bc.) Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis supervisor prof. Ing. Karel Janda, M.A., Dr., Ph.D.

# Abstract

Although biofuels have drawn the attention of researchers since its boom, which took place 20 years ago, doubts about benefits which their usage brings in the academic debate. This thesis joins the debate that discusses the impact of biofuels on food prices. The prices of 38 commodities and assets that are related to the biofuels are examined under Minimum Spanning Tree and Hierarchical Tree methods over the years between 2003-2019. The time span is divided into 4 periods, that responds to the development of world food prices. The results show that the relationship between biofuels and their feedstock depends on the overall level of food prices. In the case of higher food prices, the link between feedstock and biofuel is stronger and therefore the price transmission is more likely to happen. With lower food prices, this link is significantly weaker. Furthermore, the development of world food prices does not follow the trend of increasing biofuels production as food prices have become stable in recent periods. Therefore, this thesis does not support the claim that biofuels cause higher prices of food.

## Keywords

biofuels, ethanol, biodiesel, transportation, food price transmission, Minimum Spanning Tree, Hierarchical Tree

# Abstrakt

Přestože biopaliva přitahují pozornost vědců již od svého rozmachu, který začal před 20 lety, v akademické diskusi stále panují nejasnosti ohledně výhod jejich používání. Tato práce se připojuje k debatě, která řeší vliv biopaliv na cenu jídla. 38 časových řad cen komodit a dalších aktiv, které se vážou k biopalivům, jsou zkoumány během let 2003-2019 pomocí metod minimální kostry grafu a hierarchického stromu. Toto časové rozpětí je rozděleno do čtyř period, které odpovídají vývoji světových cen jídla. Výsledky ukazují, že vztah mezi biopalivy a jejich rostlinými složkami závisí na celkové hladině světových cen jídla. V případě, že ceny jídla jsou vysoké, vztah mezi biopalivem a jeho rostlinou složkou je silnější a proto mezi nimi může snadněji dojít k přenesení ceny. Při nižší ceně jídla je tento vztah výrazně slabší. Mimoto, vývoj světových cen jídla nekoresponduje s trendem vzrůstající produkce biopaliv. V poslední době se ceny jídla ustálily, kdežto produkce biopaliv vytrvale roste. Z těchto důvodů tato práce nepodporuje tvrzení, že biopaliva způsobují vyšší ceny jídla.

## Klíčová slova

biopaliva, ethanol, biodiesel, doprava, převod cen jídla, minimální kostra grafu, hierarchický strom

## Declaration of Authorship

I hereby proclaim that I wrote my bachelor thesis on my own under the leadership of my supervisor and that the references include all resources and literature I have used.

I grant a permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, 8 May 2019

Signature

## Acknowledgment dedication

I would like to express my gratitude to my thesis supervisor prof. Ing. Karel Janda, M.A., Dr., Ph.D. for his valuable advice and feedback throughout the process of writing the thesis. Furthermore, I would like to thank to Jana Holková for her support, help and never-ending smile.

This thesis is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 681228.

# Bachelor's Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Author's name and surname: Matěj Novotný

E-mail :matysek.no@gmail.com

Phone: 602472927

Supervisor's name: prof. Ing. Karel Janda M.A., Dr., Ph.D.

Supervisor's email: Karel-Janda@seznam.cz

*Notes: Please enter the information from the proposal to the Student Information System (SIS) and submit the proposal signed by yourself and by the supervisor to the Academic Director ("garant") of the undergraduate program.*

**Proposed Topic:**

Estimating the Relationship between Food, Fuel and Biofuel Prices

**Preliminary scope of work:**

***Research question and motivation***

In today's age of climate changes, when our civilization faces the phenomena of global warming, biofuels are widely discussed topic. The opponents of adding organic feedstock into the fossil fuels argue, that an increase in using biofuels leads to higher prices of food.
Therefore, the biofuels-related system was recently an object of scientific research. An innovative approach to this discussion appeared in the article *Correlations between biofuels and related commodities: a taxonomy perspective* (Krištoufek et al., 2012). Authors tried to estimate the relationship with the taxonomy perspective and they discovered some positive links between biofuels and feedstock commodities. However, they pointed out that there are non-negligible differences between particular types biofuels as their links to food feedstock differ quite significantly. Later on, Filip et al. (2016) built on their work with article *Dynamics and evolution of the role of biofuels in global commodity and financial markets*. They extended the number of observed commodities, prolonged the time series data and utilized the results in a different manner with wavelet analysis.

The core questions, which I should try to answer in our thesis are following:
How did the biofuels-related system prices change in the last years of fast economic growth?
Is it possible to find new commodities, that would have a significant effect on the biofuels-food system?

***Contribution***

As the debate about prices of food still persists up to today, our thesis could help understand the relationships between food feedstock and biofuels in the bigger depth. It could also determine the direction of their effects. Does the price of biofuels lead price of food or vice versa? As long as some research has been done in this area, we would like to update the datasets and figure out the behaviour of these commodities in today's fast-growing economies and possibly extend the spectrum of observed commodities as it is suggested by Krištoufek (et al., 2014). Such empirical results could contribute to the policy-making debate about biofuels.

***Methodology***

In our thesis, we would like to use 2 empirical methods, which are in the topic of biofuels still rather new, but their results have been already approved by research. In the first part, we would like to use taxonomy methodology of minimum spanning trees and hierarchical trees, which was first introduced to biofuels topic by Krištoufek et al. (2012). Here we would like to identify the most important components of the food-fuels-biofuels system and recognize where the strongest links are. For this purpose, we are going to use data from Bloomberg database, from which we can obtain time series data of various needed commodities such as biofuels, fossil fuels, feedstock or stock indices. As a result of the first empirical part, we should receive hierarchical structure of our dataset. We should be able to estimate the commodities that have a strong link to particular biofuels. On these links, we would like to build on in second empirical part, where we are going to use wavelet coherence analysis, first utilized in biofuels context by Krištoufek et al. (2014). It provides us a powerful tool, which enables us studying chosen biofuels links both in time and frequency domains.

***Outline***

1. Introduction
2. Literature Review
3. Data introduction
4. Empirical analysis
   4.1 Minimum spanning trees method
   4.2 Wavelet Analysis
5. Analysis results
6. Conclusion

**List of academic literature:**

***bliography***

FILIP, Ondrej, et al. Food versus fuel: An updated and expanded evidence. *Energy Economics*, 2017.

FILIP, Ondrej, et al. Dynamics and evolution of the role of biofuels in global commodity and financial markets. *Nature Energy*, 2016, 1.12: 16169.

FILIP, Ondřej. Food vs. Fuel: The Role of Bioenergy. 2015.

KRISTOUFEK, Ladislav; JANDA, Karel; ZILBERMAN, David. Price transmission between biofuels, fuels, and food commodities. *Biofuels, Bioproducts and Biorefining*, 2014, 8.3: 362-373.

KRISTOUFEK, Ladislav; JANDA, Karel; ZILBERMAN, David. Correlations between biofuels and related commodities: A taxonomy perspective. 2012.

SERRA, Teresa; ZILBERMAN, David. Biofuel-related price transmission literature: A review. *Energy Economics*, 2013, 37: 141-151

VÁCHA, Lukáš, et al. Time-Frequency Dynamics of Biofuel-Fuel-Food System. *Energy Economics*, 2013, 40: 1

# Contents

# Introduction

Since the beginning of the 21st century, biofuels have experienced fast development and their production grew more than four times from 33 billion liters in 2004 (Timilsina and Shrestha, 2010) to 143 billion liters in 2017 (*Renewables Global Status Report*, 2018). The original enthusiasm for a new environmental-friendly energy source, that would also solve the issue of energy insecurity for many countries, was later displaced with mixed feelings from problems that arise from the production of biofuels on the larger scale. Although the topic has become an object of investigation by many researchers, two issues that are essential for the future of biofuels are still unresolved.

First, it is the attitude of biofuels towards climate. The primary incentive to the large increase in biofuels usage was motivated by the idea, that biofuels significantly decrease the emission of carbon dioxide compared to conventional fossil fuels. However, this idea was challenged by Searchinger et al. (2008) and since then, researchers are not able to resolve this issue with the a clear definitive statement. Second, it is the position of biofuels in the agricultural price transmission system. The intensive increase in the production of biofuels changed the allocation of agricultural commodities. The land is a finite resource and farmers have to decide which crops they will grow and to whom they will sell them. Therefore, crops that are required in biofuels production (e.g corn or sugarcane) may decrease overall food production. According to some studies, this may lead to a significant increase in food prices (Pimentel et al., 2009). On the other side, despite the fact that there still exist some ambiguities in the behaviour of biofuels in the biofuels-food system, recent research supports the idea that biofuels are not among the main causes of food price increases (Taghizadeh-Hesary et al., 2019). Moreover, this statement matches with the conclusion of the review of biofuels literature delivered by Janda and Krištoufek (2019), which will be more discussed in the chapter Literature review.

This thesis joins the research that deals with the price relationship of biofuels with other agricultural commodities. The intention of the thesis is to conduct a comprehensive analysis of food-fuels-biofuels system over the time. In order to do that, three main objectives are set.

First, the thesis aims to discover the links between biofuels and related assets. For the purpose of the analysis, we gather together a unique dataset of 38 time series of prices of biofuels, food commodities, fossil fuels, exchange rates, and financial assets. For such a large amount of data, an appropriate methodological approach is required. Therefore, we utilize methods of Minimum Spanning Tree (MST) and Hierarchical Tree (HT). The methods are quite straightforward as they use simple correlations transformed into a distance metric. Nonetheless, they allow users to work with the whole network simultaneously and reveal the most important links among the network's members.

Second, the thesis seeks to examine the discovered relationships in different time periods. The dataset covers the period from 2003 to 2019, together seventeen years. During this years, biofuels underwent a rapid boom and their market changed significantly. Also food commodities experienced turbulent development. The calm period before 2007, two extreme food price hikes in years 2008 and 2012, and rather the stable period with higher overall prices after the year 2016 create various environments for links between biofuels and related commodities. Observing the links in different time periods may bring important insights into the explanation of biofuels price transmission behaviour. Third, a special focus is put on the interpretation of the explored relationships in terms of real-world events. Agricultural commodities are highly dependent on weather, financial commodities behave according to the financial cycle, crude oil is linked to the politics of its producers. The goal of the thesis is to take the information about biofuels relationships obtained during the analysis and put them into the context of the real world.

The thesis is organized as follows. The first chapter delivers Literature Review of research that described mainly the price transmission of biofuels with related commodities. A brief part is also dedicated to the research that deals with the environmental consequences of biofuels usage. The Chapter 2 provides information about the main biofuels markets. It should contextualize the reader about how large the particular market is, which feedstock is used and what policy the market utilizes. The second part of the chapter brings a short insight into the history and development of biofuels since its origin. The third chapter introduces used methodology in detail. Chapter 4 describes dataset. The individual time series are divided into groups which are discussed scrutinizingly. Stationarity and normality tests are performed in this chapter as well. The fifth chapter brings the main results of the thesis. MST and HT graphs for particular periods are presented together with an explanation of relationships between members of the network. The last part of the thesis concludes.

# Chapter 1

# Literature Review

Following chapter is dedicated to a brief review of present literature which deals with a phenomenon of biofuels and their effect on our society. Such an overview is requisite for a proper understanding of biofuels topic. The main focus is targeted on empirical research that uses a time series approach to the food-fuels system. Following structure is used. First, a short summary of literature reviewing the broad topic of biofuels is presented. Second, we would like to present some work that study links between crude oil and agricultural prices. Third, the description of literature that describes the behaviour of biofuels on the biggest world markets will be delivered. Last, a considerable part of the overview is related to research that uses proposed Minimum Spanning Tree method. Moreover, papers that brought this method to biofuels topic are discussed into slightly more detail.

The quick and progressive arrival of biofuels to the world energy market lured an extensive number of researchers who started to observe this new and promising part of the energy market. For better understanding of such a complex topic, we might find some guides that lead us through that confusing labyrinth. One of these guides is Serra and Zilberman (2013). They reviewed the main findings of time series studies that examined volatility interactions between food, biofuel, and fossil fuel markets together with delivering used methodology and datasets. Recently, Janda and Krištoufek (2019) came with well organized description of time series models in the field of biofuels.

In addition, literature with different modeling techniques is presented. With many papers reviewed, Janda and Krištoufek (2019) state that it is not possible to make a general conclusion toward biofuels price transmission as market situations differ both in time and place. Nevertheless, there is not an extensive number of research that would support the idea of biofuels driving prices of agricultural commodities up. Such a conclusion is consistent with Zilberman et al. (2012), who also summarized research regarding food-fuel transmission links.

Co-movement among crude oil and agricultural prices was the frequent target of researchers interest. Unfortunately, with the different models and data, results of the studies are frequently in the opposite meaning.

Zhang et al. (2010) discussed interactions between fuel and food commodities, both in the long- and short-run. They investigated data consisting of fuels prices (ethanol, gasoline, and oil) and agricultural commodities (corn, rice, soybeans, sugar, and wheat). The outcome suggests no long-run price relationship. The short-run one may be quite questionable too which is consistent with the work of Filip et al. (2017), who replicated the study with the newer data (until 2017). They claim that there is no strong evidence of biofuels increasing the price of food significantly. Saghaian (2010) admits that he comes with no convincing answer for his research question whether there exists some causation between oil sector and commodity prices. Despite the fact that he found a strong correlation among them, the evidence of causal links is not clear. When Granger causality test implies that crude oil prices influence prices of corn, soybeans, and wheat, on the other side VEC model with the same dataset does not prove any such transmission. Pal and Mitra (2017) tried to explain long-lasting debate on the relationship between crude oil and food supply. The monthly data of food price indices and crude oil in the period from 1990 to 2016 were examined under wavelet coherence analysis. With no association prior to 2001, they found a strong correlation starting in mid-2001 and ending during the October 2016 at long-term scale (32 weeks). It may be interesting that during the years of food crisis (2006-2008) significant correlations could be visible even in short-term scales. That could mean that world food prices follow fluctuations in crude oil prices. We may find plenty of other works that investigated crude oil - agricultural commodities relationship, for instance Fernandez-Perez et al. (2016), Myers et al. (2014) or Natanelov et al. (2011).

Sizable research observed particularly US biofuels market. Du and Hayes (2012) utilized Ordinary Least Squares method to inspect the effect of growing ethanol production on the wholesale gasoline prices. The results show quite significant and negative response of gasoline, as on the average, the price decreased by \$ 0.29 per gallon. The impact of the biofuels policies in the price transmission within the food supply was examined by Drabik et al. (2016). It is highlighted that one has to distinguish between various policies. In the situation when the binding blender′s tax credit is the reason for ethanol production, the price shock coming from food chain impacts the corn market in the lower extent than with no ethanol production present. On the other hand, that is not the case of a blend mandate, where the price transmission occurs in the same rate independently to biofuels production. Therefore, appropriate policy intervention should be implemented with a caution.

No less important ethanol producer is Brazil. Recently, Dutta (2018) studied the casual relationship between crude oil, ethanol, and sugar prices. The analysis reports several interesting outcomes. Their conclusion is in favour of the idea that sugar prices lead ethanol prices and not vice versa. The author also suggests that the sugar market in Brazil is resistant to an international oil price shocks. Hence, he assumes that an increase in the ethanol usage in Brazil secures its partial independence of the world fossil fuel market. Last but not least, the study plays an important role for policymakers. It helps to answer widely discussed question whether sugar prices are influenced by the changes in ethanol or crude oil. Quite strong evidence that sugar does not respond to fluctuations of mentioned commodities is delivered. De Gorter, Drabik, Kliauga and Timilsina (2013) studied policies affecting Brazilian biofuels market. Extraordinary conditions of traditional ethanol-using country induce unusual consequences of used fuel policies compared to other countries. Whereas in the US low gasoline tax and high anhydrous ethanol tax exemption help fuel industry, in Brazil they have got rather a harmful impact. Khanna et al. (2016) examined the distributions of impacts of Brazilian fuels policies. In other words, who benefits from them and who is reversely worse-off.

In spite the fact that we might find a sizable research of biofuels-feedstock system for individual countries, similar research at the international market level is noticeably more modest. The logical candidates for such research are two biggest ethanol producers the USA and Brazil. This research gap was recently covered by Capitani et al. (2017). They firstly observed both markets domestically and then looked into the market co-integration. In both countries ethanol is affected by international oil prices. A noteworthy difference occurs in ethanol-feedstock relationship. While in Brazil sugar prices cause ethanol prices, in the USA the trend has opposite direction, ethanol leads corn. Second, the international causality was studied which resulted into the finding that Brazilian ethanol influences the price of US ethanol. The proposed model implies that the variables of corn, sugar, oil, and Brazilian ethanol prices determine 46% of the variation in the US ethanol prices. In Brazil, ethanol market is more independent as given variables (sugar, US ethanol, and oil prices) explain only 20% of the variation in the Brazilian ethanol market.

Generally, Brazilian and US biofuels market enjoyed a broad attention of researchers as they observed them from many different angles. However, the European biofuels market stood for a long time aside of an academic debate. The main distinction from the already described markets is the type of used biofuel. Unlike US/Brazil ethanol, Europe focused on biodiesel. Predominant feedstock component in European biodiesel is rapeseed. As long as it is part of both food and biofuel supply, the biodiesel market faces the same concerns as the ethanol one. Does the recent significant growth in the biodiesel production affect agricultural commodity prices? Bentivoglio et al. (2014) take into consideration the answer to this crucial question for decisions of policymakers. They observed the price transmission between rapeseed oil and biodiesel under VEC Model. Analysis reports different results for the long-run and short-run. In long-term a positive correlation between biodiesel and rapeseed oil and diesel prices is noticeable, with the link to diesel not actually strong. The impact of diesel on biodiesel is not significant in the long-run. The authors propose an explanation that biodiesel and diesel usually work as complements because the European Union focuses on blending them together (7% biodiesel, 93% diesel). In the short-term, biodiesel does not react to changes of rapeseed oil. In conclusion, the authors argue that EU biodiesel does not have enough power to influence food prices. Hassouneh et al. (2012) obtained similar outcome, while they centered their attraction toward Spain biodiesel market. According to them, biofuels dispose of only restricted power to affect food prices. The objective set by De Gorter, Drabik and Timilsina (2013) was to show how the various biodiesel policies behave under a change in crude oil prices. The model aims to detect the distinctions between a biodiesel consumption subsidy (tax exemption) and a blending mandate. The findings show that a crude oil price shock, transmitted into higher diesel price, has completely opposite consequences while implementing different policies. On the one hand, the higher diesel prices boosted biodiesel price under a tax exemption, on the other hand, it weakened biodiesel prices under a binding blending mandate. Among other literature that is focused mainly on biodiesel it could be named Busse et al. (2010*b*), Busse et al. (2012) or Abdelradi and Serra (2015).

One of the most prominent arguments for progressive biofuels policy implemented recently by governments was a statement that biofuels could lower greenhouse gas emissions compared to traditional fossil fuels. However, this claim was in the recent years challenged by several researchers ((e.g. Plevin et al. (2010); Yang et al. (2012)). For instance, Plevin et al. (2010) questioned models that do not assume the effect of indirect land-use change which is basically the effect of changing traditional flora to the crops used for biofuels production. This effect could be significantly negative, causing higher greenhouse gas (GHG) emissions than without the usage of biofuels. However, their research could not produce any consistent results on an overall biofuels GHG stance. Another contribution to this unresolved debate with a crucial importance for the future of the whole biofuels concept brought Piroli et al. (2014). Their analysis vindicates the original green incentives of policymakers. First, the single channels through which biofuels influence CO2 emissions are explained. Among negative ones, that increase CO2, we may count indirect land use change, carbon leakage and crop yield effect. The channels that decrease CO2 emissions are fuel substitution effect and consumption effect. The proposed model under the SVAR model delivers the following results. Despite the fact that in the short-run (2-3 years) biofuels actually increase CO2 emission temporarily from various reasons (mainly because substitution effect has certain time response to biofuel production), in the long-run (starting from the fourth year) the reduction of CO2 emission caused by the usage of biofuels seems to be considerable. In such circumstances their positive environmental impact on the Earth is unquestionable. Rajcaniova et al. (2014) tried to estimate the effect of bioenergy on global land-use and for purpose of this thesis, biofuels role in this topic is essential. We may recognize two channels of land-use change. First, it is the expansion of agricultural area to the land that was not used to serve for agriproduction, called indirect land change impact. Direct land change impact is then the substitution of land from food commodities to bioenergy feedstock. Estimations reveal the presence of both types of land-use change. Agricultural area for planting biofuels crops increases each year by 0.25% of the worldwide agricultural area. Substitution effect indicates the trend, where soybean supersedes grassland and rice land.

In the last part of the literature overview, we would like to discuss an unorthodox branch of research that brought quite innovative methods to the world of biofuels. From the perspective of this thesis, Krištoufek et al.

(2012) stand in the centre of attention as they proposed for the first time Minimum Spanning Tree (MST) and Hierarchical Tree (HT) methods to the biofuels-related topic. Naturally, this kind of methodology has got several advantages (mainly enable us to work with the immense number of variables in the dataset), as well as some disadvantages (does not allow us to study causality between variables). The authors utilized broad dataset of biofuels, fossil fuels, crude oil, and agricultural commodities. The main attention was paid obviously to the relationship of biofuels (ethanol and biodiesel) with other commodities. Moreover, for better interpretation,the time span of the analysis was split to the pre-crisis period (2003-2007) and post crisis-period (2007-2011). The results prove that this decision was well-founded, because both periods show different links. For the first period, it holds that neither biofuels nor agricommodities (soybeans, wheat, and corn) are correlated and therefore connected to the rest of the system. The researchers conclude, that in the situation when food prices are low, there exists only a weak link from the food commodities to the fuels and biofuels. Different findings hold for the post-crisis period of higher food prices. In this case, we have to distinguish carefully between ethanol and biodiesel. While ethanol is quite strongly related to corn, wheat, and soybeans, biodiesel is more correlated with fuel branch in the commodity network. Corn, wheat, and soybeans are strongly connected with the overall system. Therefore, in the conclusion it is emphasized that it is necessary to distinguish between particular types of biofuels. Vacha et al. (2013) used wavelet coherence methodology for the first time at the field of biofuels and such approach allow them to study the relationship between commodities both in time and frequency. It is showed, that at the market, two major links could be observed: ethanol and corn link, and biodiesel and German diesel link. These relationships are strongest at long-term on the frequency of 32 weeks and their direction goes from the corn to ethanol and form German diesel to biodiesel. However, the strength of the causation changes over the time and it appears that with lower prices of corn (diesel), its leadership is more significant. The similar methodological procedure was presented by Krištoufek et al. (2016) while they focused solely on ethanol-related prices. MST and wavelet coherence method were brought together in the work of Filip et al. (2016). First, they discovered the most prominent links between used variables (33 time series of prices of biofuels related commodities and assets) via MST analysis. Subsequently, explored links were analyzed under wavelet coherence method. As an outcome, they received the important links with the information which commodity leads another one. Additionally, they controlled for the impact of crude oil in these links. The conclusion of the paper converges to the results of the previous authors. The prices of US and Brazil ethanol depend on the prices of their feedstock. The reverse influence is not visible. European biodiesel stands outside of establishment of the traditional ethanol markets while it does not respond to the prices of its feedstock significantly. For biodiesel, the price of fossil fuels is much more important.

# Chapter 2

# Biofuels overview

## 2.1 Market description

The global market with biofuels has not changed dramatically since its origin at the beginning of the century. The main market share is still covered by three biofuels giants - the USA, Brazil and the European Union. Nevertheless, over the years we could have witnessed miscellaneous policy implementations, their subsequent modification or complete changes in the direction of the biofuels approach. Countries′ production and consumption quantities went under certain evolution as well. For that reasons market and policy situation of the three biggest players are described in this chapter.

### 2.1.1 The global market

At the beginning of the 21th century, conventional biofuels experienced an extraordinary boom. According to *Statistical Review of World Energy* (2018), between the years 2006-2016 the annual growth in production reached 11.4%. The main phase of the rapid growth ended in the year 2014. Since then, the fast growth has slackened and in 2017 we witnessed gradual 2.5% production growth. The overall production in 2017 attained 143 billion liters (*Renewables Global Status Report*, 2018). Decomposing this number, the study estimates 105.5 billion liters in the ethanol production and 36.6 billion liters in the biodiesel production. The rest of production belongs to minor biofuels like hydrogenated vegetable oil. The production and consumption of biofuels are highly associated with a geographical location. More than 80% of biofuels are produced and consumed in the USA, Brazil, and the EU.

### 2.1.2 The USA

In the United States, biofuel industry produced in 2017 around 61 bln L of ethanol and 6.9 bln L of biodiesel which makes the USA the largest biofuel producer in the world (FAO, 2018). According to the study, the USA should maintain its biofuels primacy in a near future. It predicts very similar production output in 2027 compared to 2017. In recent years, the ethanol production grew decently as the consequence of higher targets of the Renewable Fuel Standard (RFS).

Dominant ethanol feedstock on the US market is corn. Geographical conditions do not allow greater sugarcane production, which is more efficient ethanol feedstock. Biodiesel is mainly produced from soybean, followed by corn and canola oil (EIA, 2019)

The US biofuels policy is included in the RFS, which was established in year 2005 and rescheduled in year 2007. It sets quantitative mandates of biofuels for each year. According to this plan, the biofuels production should increase from 12.95 bln gallons in 2010 to 36 bln gallons in the year 2022 (Bracmort, 2018). In reality, the blending share of biofuels in gasoline exceeded 10% in 2017 (*Renewables Global Status Report*, 2018).

### 2.1.3 Brazil

Brazil is the second most important biofuel producer. The major focus of Brazilian biofuels programme is orientated on ethanol due to a good availability of sugarcane, which is the easiest processable ethanol feedstock. Brazil production was 28.3 bln L in the year 2017 and estimates indicate a raise of 9% in the following year (FAS, 2018*a*). Biodiesel plays only the secondary role at Brazilian biofuels market. Nevertheless, changes in biofuels policy help biodiesel to strengthen its position. In 2017, it accounted for 4.3 bln L, in 2018 it is expected an expansion to 5.4 bln L. Majority of biodiesel is prepared from soybean (FAS, 2018*a*). By far the most important ethanol feedstock is sugarcane as it has no serious market competitor. Ethanol takes approximately 61% of total sugarcane plants in Brazil. The number recently increased due to the surplus of sugar in the world market which weakens its price (FAS, 2018*a*).

Brazil has a long-standing tradition in high ethanol blending mandates. In 2015, government enforced the 27% mandate, but already in 2006 there was the 25% minimum blending limit. For the high demand for ethanol there is not sufficient domestic supply if the high prices of sugar do not support sufficient conversion of sugar cane into ethanol. Therefore, depending on sugar cane productions and sugar market prices Brazil may have to import missing ethanol sometimes, primarily from the United States. More significant policy adjustments undertook the biodiesel market. In March 2017, 8% blending mandate took effect. Only a year later policymakers decided to augment the limit to 10% (FAS, 2018*a*).

### 2.1.4 European Union

The European Union is the biggest producer of biodiesel in world with 13,5 billion liters (37% of the world production) (FAS, 2018*b*). The leaders of European production are Germany, France, and Spain. However, FAO (2018) predicts a decrease of biofuels production, which should in 2027 account for 12.9 bln L. Such trend could be related with cheap biodiesel that is imported from Argentina and Indonesia after removal of anti-dumping duties for these countries by European commission (FAO, 2018). Besides biodiesel, member states of the EU generate also ethanol, mainly from grains and sugar beet derivatives. Its production attained 5.3 bln L in 2017.

The main biodiesel feedstock in the EU represents rapeseed oil with 45% of production in 2017. Nonetheless, its usage is continually falling as in 2008 it accounted for 72%. Used cooking oil is the second most used feedstock receiving 21% and palm oil occupies the third position with 18% of total feedstock (FAS, 2018*b*).

European policy toward biofuels is determined by the Renewable Energy Directive (RED) which sets requirements in renewable resources politics in the period 2009-2020. It also includes a target for the transport sector. The blending mandate for biofuels should fulfill a limit of 10% in the year 2020. In the reality, this goal is not likely to be accomplished. In the year 2015, blending mandates reached only 6% share and only Sweden and Finland exceeded the limit. The European Union has already prepared an updated renewable policy for the years 2021-2030 under the name the RED II, which counts with 14% of renewable energy in the transport sector by the end of 2030 (FAS, 2018*c*).

## 2.2 Development of biofuels

While biofuels do not play the most fundamental role in today′s energy market, some may be surprised how rich and gripping history they have. Therefore, in this section summary of the most important highlights in the biofuels development is presented. Singh (2013) serves well as a great source of biofuels knowledge and this part is based on information gathered from it.

The first liquid fuels developed by mankind were biofuels such as vegetable oils, animal fats, or ethanol from crops. They were a relevant part of the beginning of industrial revolution, inventors used them in lamps and in internal combustion engines. The most widespread biofuel was camphane which was an irregular blend of ethanol, turpentine and camphor oil. However, since 1860s biofuels started to lose the market battle with petroleum products like kerosene and gasoline. In Europe, the trend was moderate while in the USA imposed taxes on alcohol caused a radical shift to petroleum. With no serious market competitor, petroleum products became unwavering leaders of the fuel market. The situation was not influenced even by the repeal of the tax on biofuels in 1906, although the US automotive industry enforced usage of biofuels. A big proponent of biofuels was Henry Ford, whose famous Model T was constructed for either gasoline or alcohol. Biofuels position in Europe was slightly different due to no alcohol tax and only poor domestic oil reserves. The biggest support to biofuels was given by Germany whose small number of colonies provided almost no oil supply. Germans invested a big effort to potato alcohol, which was seen as the future substitute for petroleum.

Between the wars, scientists slowly became aware of the non-renewability of fossil fuels, but they failed to communicate it to the energy market, which paid to biofuels only a negligible attention. A broad attention biofuels received firstly during the oil shock in the 1970s. The shortage in supply of cheap oil from the Middle East induced a deep energy crisis. After this experience, developed economies started to perceive biofuels as a great opportunity how to decrease their dependence on imports of oil. Brazil quickly realized the benefit of its abundant production of sugar, which may be easily transformed to ethanol, and introduced the mandatory blending of 20% of ethanol into gasoline. Similarly, the USA implemented ethanol programme in 1980. It included tax exemption on ethanol and prohibition of Latin American ethanol on the market.

The fundamental incentive for boosting biofuels production came during the debate about increasing the octane number in gasoline. Basically, the octane number means efficiency in gasoline combustion. Higher it is, more efficient the burning of gasoline in an engine is. Although the luck of biofuels lies in its characteristics that they improve the octane number of gasoline when they are blended together, historically, there existed cheaper and more convenient additives for petroleum industry that achieved the higher octane number. First of all, it was a lead-based additive - the tetraethyl lead. Despite the fact, that there were various concerns about the impact of leaded gasoline on the public health since its beginning, lead was used in gasoline until the end of 20th century. Scientific studies proved that lead exposure causes cardiovascular diseases, hypertension, and other fatal illnesses (*A Brief History of Octane in Gasoline: From Lead to Ethanol*, 2016). At the end of 20th-century petrol industry started to use a new additive, which could raise octan number - the methyl tertiary butyl ether (MTBE). The MTBE quickly expanded to become the major gasoline additive. However, the wide usage of the MTBE raised worries about its effect on the environment.In the year 2000 the phase-out of the MTBE was announced when the proofs confirmed it as a water pollutant and a possible carcinogen (Krištoufek et al., 2012). Raising global concerns about the climate change, the GHG emissions together with the gap in the octan number additives market created the ideal conditions for the biofuels revival. Biofuels undertook the position of the MTBE and became the most important additive of today. Their undisputed market advantage consists of government support via various policies, because it sees biofuels as an influential tool in the climate change fight.

# Chapter 3

# Methodology

The present chapter introduces the empirical methods used in the thesis. As mentioned in the Introduction, the empirical research is based on minimum spanning trees (MST) and hierarchical trees (HT). Therefore, the mechanism behind these methods will be described in a detail. These, now, time-tested methods were firstly brought to financial time series data by Mantegna (1999), who was quickly followed by other researchers - Bonanno et al. (2004), Tabak et al. (2010), or Lautier and Raynaud (2012) for energy derivatives. Krištoufek et al. (2012) was then first who applied the approach in the biofuels perspective. Listed authors also serve as the key source for information presented in the subsequent methodological parts.

## 3.1 Distance metric

The sample correlation coefficient is a common practice in measuring the relationship between time series. For a pair of assets $i$ and $j$ with values $X_{it}$ and $X_{jt}$ and $t = 1, \ldots, T$, the sample correlation coefficient $\widehat{\rho_{\mathrm{ij}}}$ is defined as

$$\widehat{\rho_{ij}} = \frac{\sum_{t=1}^{T}(X_{it} - \bar{X}_i)(X_{jt} - \bar{X}_j)}{\sqrt{\sum_{t=1}^{T}(X_{it} - \bar{X}_i)^2 \sum_{t=1}^{T}(X_{jt} - \bar{X}_j)^2}} \tag{3.1}$$

where $\bar{X}_i$ and $\bar{X}_j$ represent the time averages, which are calculates as $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$ and $\bar{X}_j = \frac{1}{T}\sum_{t=1}^{T} X_{jt}$. By the definition, values of the $\rho_{ij}$ can vary in the interval $[-1, 1]$, where -1 stands for the perfect negative correlation, 0 stands for no correlation and 1 stands for the perfect positive correlation. Such correlation coefficient has a meaning only for the time series with well defined means and variances. Therefore, our time series have to satisfy the (weak) stationarity condition (Krištoufek et al., 2012).

According to Verbeek (2008) a stochastic process is stationary if its properties do not change over time. In this case it is sufficient to define the weak stationarity which requires only the first and the second order moment of a process to be constant.

**Definition 1.** *A process $Y_t$ is defined to be weakly stationary if for all $t$ it holds that*

- $E\{Y_t\} = \mu < \infty$
- $V\{Y_t\} = E\{(Y_t - \mu)^2\} = \gamma_0 < \infty$
- $Cov\{Y_t, Y_{t-k}\} = E\{(Y_t - \mu)(Y_{t-k} - \mu) = \gamma_k, \qquad k = 1, 2, 3, \ldots$

In this paper we consider a dataset containing N commodity prices in the time perspective. Observing every single connection between all commodities generates $N\ (N - 1)/2$ correlations, which can be written in the form of matrix $\mathbb{C}$. The individual elements of $\mathbb{C}$ are defined as:

$$\rho_{ij} = \begin{cases} \text{corr } (x_i, x_j) & \text{if i} \neq \text{j} \\ 1 & \text{if i} = \text{j} \end{cases} \tag{3.2}$$

Matrix $\mathbb{C}$ has several important properties. First, $\mathbb{C}$ is a squared matrix of order n. Second, since there is no difference in the order of $i$ and $j$ in the correlation coefficients, $\rho_{ij}$ and $\rho_{ji}$ are the same, the matrix is also symmetric. Third, all elements on diagonal are equalled to 1, since the correlation between the same variable is the perfect correlation.

In order to define MST, a variable that might be interpreted as the distance is required. The correlation matrix designed in formula (3.2) therefore cannot be employed as it violates the three axioms of Euclidean metric. However, according to Mantegna (1999), there exists a non-linear tranformation $d_{ij}$ that transforms correlation $\rho_{ij}$ into a distance measure:

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}. \tag{3.3}$$

With $d_{ij}$ Euclidean metric is satisfied:

1. *Identity*: $d_{ij} = 0 \iff i = j, \forall i, j \in \mathbb{N}$
2. *Symmetry*: $d_{ij} = d_{ji}, \forall i, j \in \mathbb{N}$
3. *Triangle inequality*: $d_{ij} \leq d_{ik} + d_{kj}, \forall i, j \in \mathbb{N}$

Values received from the non-linear transformation $d_{ij}$ are on the contrary to coefficients $\rho_{ij}$ strictly positive and range from 0 to 2, where $d_{ij} = 0$ is the perfect positive correlation and $d_{ij} = 2$ is the perfect negative correlation. No correlation corresponds to the value $d_{ij} = \sqrt{2}$

Similarly as for correlation coefficients $\rho_{ij}$, we may also compound the distance matrix $\mathbb{D}$ for distance coefficients $d_{ij}$ . The individual elements of $\mathbb{D}$ are defined as:

$$d_{ij} = \begin{cases} \sqrt{2(1 - \rho_{ij})} & \text{if i} \neq \text{j} \\ 0 & \text{if i} = \text{j} \end{cases} \tag{3.4}$$

Distance matrix $\mathbb{D}$ gathers every possible connection in the graph network.

## 3.2 Minimum Spanning Tree

Before we come to definition of MST, it is convenient to make a brief introduction to the graph theory. The following definitions are taken from

Nešetřil et al. (2009) and Diestel (2012), who offer a deeper source of graph theory knowledge.

**Definition 2.** *A graph G is an ordered pair (V,E), where V is some set and E is a set of 2-point subsets of V . The elements of the set V are called vertices (also nodes or points) of the graph G and the elements of E edges (also lines or links) of G.*

**Definition 3.** *We say that a graph G is connected if for any two vertices* $x, y \in V(G)$*, G contains a path from x to y.*

**Definition 4.** *A tree is a connected graph containing no cycle.*

**Definition 5.** *Let* $G = (V, E)$ *be a graph. An arbitrary tree of the form (V,E'), where* $E' \subseteq E$*, is called a spanning tree of the graph G. So a spanning tree is a subgraph of G that is a tree and contains all vertices of G.*

Basically, a spanning tree is any tree that connects all vertices. However, there may exist many spanning trees that connect the vertices in a different manner. The idea behind Minimum Spanning Tree lies in choosing the most important edges between vertices. However, the term " *the most important edges*" is quite vague and nonspecific. In order to be able to distinguish the importance of particular edges, we need to assign to every edge a numerical value called weight. In a more formal way, for every edge $e \in E$ there exists a number $w(e)$. According to these usually nonnegative weights, it may then be decided on the importance of the links.

In the case of this work, it means that we would like to choose the spanning tree in a such way that sum of the total distance between the commodities is minimal. As a result, from the original $N(N-1)/2$ connections between all assets, only $N-1$ most important with the smallest possible total distance will be picked.

As a solution to how exactly the MST should be selected, we may use several algorithms. Two most widespread are *Kruskal's algorithm* and *Prim's algorithm*. This thesis will follow the work of Krištoufek et al. (2012) and

Filip et al. (2016) and utilize *Kruskal's algorithm* as explained in Kruskal (1956).

The Kruskal algorithm works as follows: The algorithm takes all $N(N-1)/2$ connections and orders them into non decreasing sequence by the value of their weights. Subsequently, the algorithm picks the edge with the lowest value, i.e. the most important one. Then it takes the edges one by one in order of the non decreasing sequence from the lowest weight. If one or both vertices of the chosen edge do not occur in the network, the edge is added. In the situation when both vertices that are connected by a chosen edge already occur in the system, the algorithm does not choose the edge because it would create an undesirable loop. As a result, the algorithm creates a network with all given vertices and with $N-1$ connections that have the minimal total weight. Such a network is called Minimum Spanning Tree.

## 3.3 Hierarchical Tree

In addition to MST, which may provide us with rather geometrical information about individual assets and their position in the network, the thesis also employs Hierarchical Trees, which may help us in discovering the behaviour of the whole clusters of the assets and in showing the taxonomy perspective of the dataset. HT gives also an idea about the strength of relationships in a cluster.

Technical construction of HT is based on MST with one additional condition that is the definition of *ultrametric distance* $d^*_{ij}$, which not only satisfies all three axioms of Euclidian metric but also the *ultrametric inequality*:

$$d_{ij} \leq \max(d_{ik}, d_{kj}), \forall i, j \in \mathbb{N}. \tag{5.5}$$

The *ultrametric distance* $d^*_{ij}$ may be written in matrix notation as *the subdominant ultrametric distance matrix* $\mathbb{D}^*$. Individual *subdominant ultrametric distances* $d^*_{ij}$ are obtained as a maximum Euclidean distance that has to be crossed between two successive vertices during the shortest path from vertex $i$ to vertex $j$. More formally:

$$d_{ij}^{*} = \max d_{kl} \tag{3.5}$$

where $k$ and $l$ are any vertices on the path between $i$ and $j$, included. For every MST there exists the same number of subdominant ultrametric distances as the number of edges in the network, that means $N - 1$. As every single $d_{ij}^{*}$ in $\mathbb{D}^{*}$ is observed, the construction of Hierarchical Tree may begin. The process is similar to MST. The minimal $d_{ij}^{*}$ is chosen to establish the first pair of the tree and subsequently from the smallest to the greatest value of $d_{ij}^{*}$ additional assets are connected to the system. Again, it holds the rule that no pair which is already in the network can be connected with the new link. The tree is complete when the last asset joins the system.

# Chapter 4

# Data

Besides the empirical analysis itself, it is the choice of the appropriate dataset which is maybe even more crucial for delivering consistent and precise results. Therefore, this chapter discusses mainly the process of data gathering and the characteristics of the collected dataset. For the purpose of this thesis, we gathered together 38 price time series of various commodities and assets from various sources while the majority of the time series were taken from Thomspon Reuters Eikon.

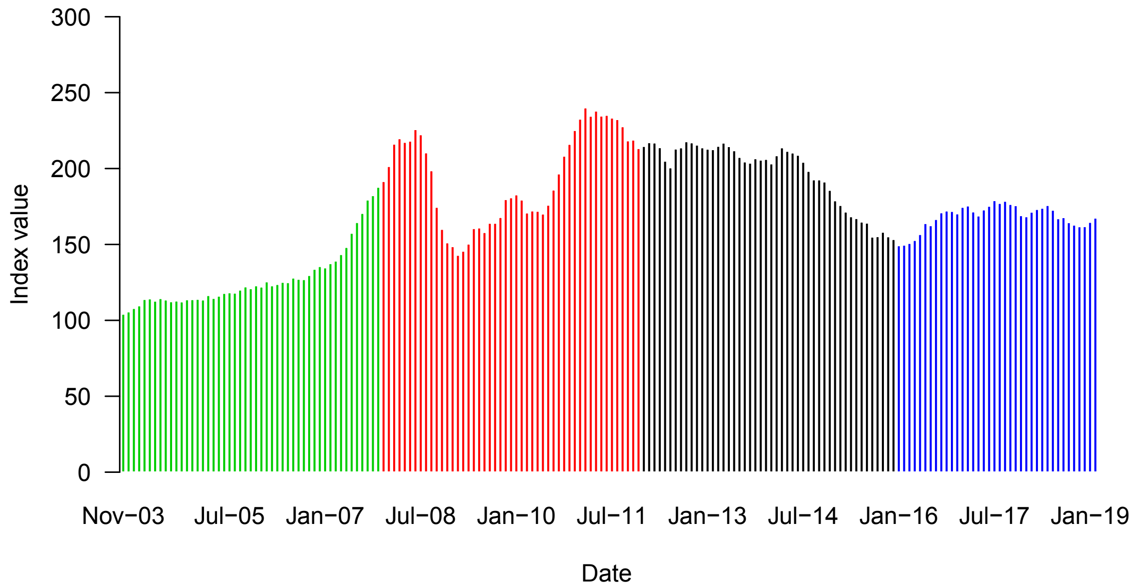
## 4.1 Data Overview

Every time series refers to weekly prices in the period from 21st November 2003 until 22nd February 2019. The day of the collection is Friday. In the situation, when the Friday data are not available for some reasons, data from the first previous work day are collected. The table with all assets, their source and the abbreviation may be found in the Appendix A together with the descriptive statistics of all examined variables. Corresponding to their characteristics, the assets and the commodities are divided into individual groups which will be discussed precisely in the following section.

The analyzed dataset brings information about the price development of observed commodities and assets over 17 years, altogether 797 observations. When we inspect the data in the detail, we may notice two periods of substantially higher food prices. These periods are 2007-2008 and 2010-2012 and they refer to the two world food price crises. The reasons that stand behind these events are still not completely clear as researchers mention various ideas including natural disasters, biofuels, or financial speculation (Tadasse et al., 2016). Nevertheless, the world food crises had an important impact on political development in the recent time. Perez and Wire (2013) may serve as an example as they claim that rising food prices were one of the causes of Arab Spring because the countries involved in the unrest are heavily depended on the import of food. The spikes in the food prices may change the overall relationship between the commodities and assets. Knowing this, it may be interesting to observe the evolution of these relationship over the time. Our long enough dataset may be an appropriate choice for a such analysis. In order to find the correct periods to which we will divide the time series, we use the Food Price Index published by Food and Agricultural Organization of the United Nations (FAO). The FAO Food Price Index (FFPI), plotted in Figure 4.1, records the monthly development in 73 prices series of 23 food commodities which are split into five price indices groups. The FFPI is measured in points in its own scale and the years 2002-2004 are taken as the base period. The plot of the FFPI shows anticipated, two hikes of the prices in 2008 and 2011. After the first hike, prices seemed to return almost on pre-crisis values in 2009, but then quickly increased on the historic maximum in winter 2011. After this peak, the FFPI constantly decreased until the end of 2015 where it hits its local minimum. Since then, the FFPI is more or less stable with the slight, gradual growth.

The FFPI implies a logical structure for our analysis, as it naturally divides our data into four subperiods. The first period represents the pre-crisis state of the market. The second period includes two prices spikes and generally represents dramatic change of the situation in the commodity market. The third period runs from the beginning of 2012, where the food market partly calmed down, until the end of 2015, where prices started once again to rise. The last period goes since the beginning of 2016 up to now. Detailed dates of the subperiod are following:

Figure 4.1: FAO Food Price Index




Source: Author's computation

1. Period 1: 21.11.2003 - 28.12.2007, 215 observations

2. Period 2: 4.1.2008 - 30.12.2011, 209 observations

3. Period 3: 6.1.2012 - 25.12.2015, 208 observations

4. Period 4: 1.1.2016 - 22.2.2019, 165 observations

The divison of the dataset in this manner should catch the changes on the biofuels market. As it was mentioned in the second chapter Market Description, the main policies affecting biofuels in the major markets were somehow reconsidered after the first price hike in 2007. The changes in the prices of the biofuels feedstock affects the situation on the market as well, which may be projected into different connections in the biofuels-fuels-foods system. Therefore, during the taxonomy analysis, we will focus on the development of the relationships in the data.

### 4.1.1 Biofuels

Since the main focus of the thesis is to find the relationship between biofuels and other assets, biofuels are in the centre of attention. As it is described earlier in the text, the biofuels markets differ significantly according to the geographical location. Hence, in order to observe the complete picture of the biofuels environment, we combine several biofuels data. The biggest ethanol producer, the USA, is represented by New York Harbor Price Ethanol index from the database Bloomberg Datastream. It provides the spot prices of anhydrous ethanol for the US market quoted in USD per gallon. Information about the second biggest ethanol producer, Brazil, are provided by CEPEA[1], an economic research centre at the University of Sao Paulo. Anhydrous ethanol index is quoted in USD per liter. Working with ethanol prices from these two major producers will ensure more than 80% share of the worldwide production. For biodiesel, two time series are employed too. Both are biodiesel (RME, spot price) produced from rapeseed (RME stands for rapeseed methyl ester) and come from Thomspon Reuters Eikon database. The time series that gives prices for US market is quoted in USD for a tonne, European one is quoted in EUR for a tonne. With commodities for these two countries, we can account for nearly 60% of the world production of biodiesel.

### 4.1.2 Biofuels feedstock

The recent debate examined the effect of biofuels on food prices. Therefore, it seems logical that our dataset should contain prices of the food which is directly involved in the biofuels production. We might see diverse number of biofuel feedstock used in the production across the world. In the case of ethanol production, the content of the sugar in the crop is essential (EIA, 2018). Simplistically, the feedstock is firstly converted into glucose. Then, yeast or other enzymes are added, and the process of fermentation transforms sugar into ethyl alcohol (ethanol), which is in the end distilled and

[1] https://www.cepea.esalq.usp.br/en

dehydrated. Plants based on sugars (sugar cane, sugarbeet) are decomposed more easily into glucose than crops based on starch (corn, wheat, grains). Hence, they are the ideal candidates for an ethanol feedstock. Thanks to the abundance of sugar cane, Brazil was able to reach by far the highest share of the ethanol usage on the domestic market in the world. However, sugar cane requires highly demanding growing conditions. For that reason, they cannot be grown everywhere (Banschbach and Letovsky, 2010). For instance, the USA, the biggest ethanol producer in the world, disposes with the limited area where sugar cane can be grown. Consequently, the most important ethanol feedstock in the USA is corn, which has much lower demands on growing conditions. The heart of the corn production is upper Midwest, sometimes called the Corn-belt (Green et al., 2018).

In the dataset, three sugar price indices are employed. To observe sugar cane behaviour on the world market, we use the Intercontinental Exchange (ICE) prices of raw centrifugal cane sugar based on 96 degrees average polarization. Sugar beets are represented by the LIFFE Sugar beets price index obtained at Bloomberg Datastream database. The extraordinary Brazilian market with sugar is approximated by the CEPEA Crystal sugar price index[2]. Starch plants have two representatives. The most important starch is corn which is represented by Chicago Board of Trade (CBOT) Corn Composite[3]. Wheat plays a minor part in the ethanol production and we quote price of No.2 Soft red winter type, traded at CBOT. Biodiesel is usually produced from straight vegetable oils or already used oils. The European Union, as the biggest producer of biodiesel in the world, relies mainly on rapeseed oil (almost 50% share). We work with rapeseed prices traded at LIFFE Paris, which is a major stock exchange for rapessed in the world (Busse et al., 2010*a*). Another important biodiesel producer, the USA, uses for its production mainly soybeans, which are represented by CBOT Composite Soybeans prices. The data for biodiesel feedstock are completed by time series prices of palm oil (CBOT) and sunflower seeds (Johannesburg

[2]price index available at www.cepea.esalq.usp.br/en

[3]the quality is marked as The second grade of yellow corn

Stock Exchange Sunflower Seed Commodity).

### 4.1.3 Fossil Fuels

Fossil fuels are a nonnegligible component of biofuels-related network. Both crude oil and conventional fuels directly affect biofuels. For the crude oil, we analyze Brent crude oil as a specification benchmark for Atlantic basin crude oils (it is extracted from the North Sea). The second used grade of the crude oil is West Texas Intermediate (WTI), traditional crude oil refined in the Midwest and Gulf Coast regions of the USA. Both types are sweet light crude oil which means they have low density and lower volume of hydrogen sulfide. Regarding conventional fuels, we obtained prices for gasoline and diesel in every major biofuel market in order to distinguish for local differences (the EU prices were estimated by German prices). The US data come from the US Energy Information Administration (EIA), the Brazilian data were downloaded from The Brazilian National Agency of Petroleum, Natural Gas and Biofuels[4] (ANP) and the German data were obtained through Thomspon Reuters Datastream. Prices for the US market were originally in USD per gallon, prices from other markets were converted from original currencies and units to USD per gallon by our own calculations.

We also enriched the system with two substitutes of gasoline and diesel. The first one is natural gas quoted at New York Mercantile Exchange (NYMEX) as Henry Hub Natural in US dollars per MMBTU which is thousand of British termal, unit a traditional unit of energy. The second one is heating oil traded also at NYMEX in US dollars per gallon.

### 4.1.4 Food

In their work, Krištoufek et al. (2012) suggest to utilize a broader spectrum of assets and commodities. In food perspective, we have already utilized a few commodities, that are directly involved in the biofuels production. To investigate possible effects that biofuels may have on the food sector, it

[4] Agencia Nacional do Petroleo, Gas Natural e Biocombustiveis - www.anp.gov.br

may be beneficial to employ some food prices that stay apart of the biofuels production. Food commodities and biofuel feedstock may compete for land, and studying the connections among the commodities over the time with the share of biofuels increasing may bring an interesting insight into the issue. Filip et al. (2016) proposed four food commodities which are also used in this thesis: rice, coffee, cocoa and oranges[5]. Furthermore, we added feeder cattle, which are calves raised to 600-800 pounds, and US cotton. The price for cattle is quoted in US cents per pound.

### 4.1.5 Interest rates

To get some macroeconomic perspective about the behaviour of global markets, two specific time series of interest rated are added into the dataset. First, it is London Inter-bank Offered Rate (LIBOR). It represents globally accepted benchmark for the interest rate which is given from one bank to the another in international market for short-term loans. LIBOR has several maturities and we used the most common one with the maturity of 3 months. The data come from the website of Federal Reserve Bank of St. Louis[6], that publishes the data on the weekly basis.

Second, the Federal funds, sometimes called Fed funds, are added into the network. The Federal reserve bank requires commercial banks to have at least 10% of the deposit they hold at the end of the day. Effective federal funds rate refers to the interest rate at which commercial banks with surplus of cash lend to other commercial banks that need to increase liquidity immediately in order to meet the FED criteria. Fed funds were also obtained from Federal Reserve Bank of St. Louis at a weekly frequency.

### 4.1.6 Stock Indices

The overall performance of the local economy may have a considerable impact on the biofuels environment. To control for this element, it could be beneficial to comprise GDP of the observed markets to the dataset. Unfor-

[5] approximated by orange juice due to the data unavailability

[6] www.fred.stlouisfed.or

tunately, the data are not available in the desired frequency - i.e. weekly. Therefore, our intention is to find a suitable proxy. For this purpose, the feasible candidates may be stock indices in the particular countries. For the USA, we adopt two indices - Dow Jones Industrial Average Index (DOW JONES) and Standard and Poor's 500 (SP 500). Europe is covered by British Financial Times Stock Exchange 100 (FTSE 100) and German Deutsche Boerse DAX Index (DAX). Further, the Brazil stock market represents Sao Paulo SE Bovespa Index (Bovespa). All data were acquired from Thomspon Reuters Eikon database.

### 4.1.7 Exchange rates

The last group the dataset is represented by exchange rates which may give us information about the mutual relationship among observed economies. The predominant currency for the most of the commodities and assets is US Dollar, so we use exchange rates of it with currencies of two other important markets - USD/EUR for Europe and BLR/USD for Brasil. Both exchange rates were obtained at Federal Reserve Bank of St. Louis.

## 4.2 Logarithmic returns

In the time series analysis of commodity and asset prices, it is standard practice to transform the prices into logarithmic return. This approach is used by the extensive number of research papers ( e.g. Sieczka and Hołyst (2009), Vacha and Barunik (2012), or Onnela (2002)). Moreover, researchers that are focused on the studying biofuels using MST, Krištoufek et al. (2012) and Filip et al. (2016), utilized in their analysis logarithmic returns. Therefore, we will also follow the procedure and transform dataset into logarithmic returns which are defined as:

$$r_t = log(P_t) - log(P_{t-1}) = log(\frac{P_t}{P_{t-1}}). \tag{4.1}$$

The usefulness of this approach will be verified in the next section. Whereas many of observed price time series are non-stationary, which would dramatically complicated our analysis, all logarithmic returns are well-behaved in terms of stationarity.

## 4.3 Stationarity

The essential property for a meaningful analysis of the time series data is stationarity. The absence of the stationarity may imply that a model that describes the data may produce results that are correct only for some time periods. In other words, the accuracy of the model may range for various time periods of the data. That is the reason why it is necessary to include tests that can detect possibility of non-stationarity in time series data. There are many tests that identify non-stationary data. In this thesis, two following tests are used.

### 4.3.1 Augmented Dickey-Fuller test

The fundamental test for the detection of non-stationarity is Augmented Dickey-Fuller test, which first version was introduced by Dickey and Fuller (1979). The null hypothesis of the test is the presence of the unit root which implies non-stationarity of the time series. The alternative indicates stationarity which one would like to prove.

### 4.3.2 Kwiatkowski-Phillips-Schmidt-Shin test

After the publication of the Dickey-Fuller tests, one might have witnessed a vivid academic debate about its ability of rejection of a unit root for economic time series (for instance Harris (1992)). To avoid the discussion about the low power of the Augmented Dickey-Fuller test, we also employ the second test for stationarity proposed by Kwiatkowski et al. (1992). This test has a different structure and tests the null hypothesis of trend stationary process against the alternative of unit root. The performance of both tests should ensure a clear result on stationarity of the data.

The outcome of these two tests may explain one of the reasons why the logarithmic returns are used instead of simple logarithmic prices. As described above, stationarity is a property convenient for time series analysis. However, when the tests for stationarity are applied to the logarithmic prices, both ADF and KPSS tests deliver undesirable results. While ADF test does not provide enough evidence to reject the null hypothesis of unit root, KPSS test rejects the null hypothesis of trend stationarity. Consequently, the test are applied on logarithmic returns and both test indicate stationarity at 95% confidence interval for every single price time series. The detailed results of the test are enclosed in the Appendix B.

## 4.4 Normality

The core of the empirical part of the text is built on the Pearson correlation coefficient which is later transformed into a distance metric. For that reason, we should not forget about an important assumption on which is the coefficient defined - normality of random variables for which the coefficient is counted (Asuero et al., 2006). In order to check normality of used data two tests are proposed. The Shapiro-Wilk test was introduced by Shapiro and Wilk (1965). The null hypothesis says that the population is normally distributed. The more the result of the test ranges from one, the stronger is the rejection of the null hypothesis in favour of the alternative, non-normal distribution of the population. The Jarque$-$ Bera test was developed by Jarque and Bera (1980). It checks whether the population has the skewness and the kurtosis of the normal distribution which is symmetric around its mean (skewness is zero) and the kurtosis is equalled to three. The value of the test is always non-negative and the smaller it is, the more certainly we reject the null hypothesis of normality of the distribution.

The two described tests for normality were applied to the data. As a result, all variables under logarithmic returns were found to have a non-normal distribution by both tests test. For logarithmic prices, the Jarque-Bera test indicates non-normality with the exception for sugar cane and BR sugar. The Shapiro-Wilk test rejects normality with no exception. The mutual rejection of normality of the data by both test means the violation of the classic assumptions of the Pearson correlation coefficient which may have negative consequences for the analysis. Detailed results of the tests may be found in the Appendix B.

Fortunately, Havlicek and Peterson (1976) came with the paper discussing the effects of a violation of the normality assumption on the Pearson correlation. The authors claim that the ineligibility to meet the distribution requirement of the Pearson coefficient affects it only a little. To quote verbatim from the study: "*It appears that the Pearson coefficient can be used in nearly all situations in which there is need for a measure of the relationship between two variables regardless of the shape of the distributions of scores*" .

After the explanation of the stationarity and non-normality issues, we may proceed to the empirical body of the thesis. Despite the fact that non-normality of the data should not create any doubts about the precision of the analysis outcomes, we should acknowledge this in the case of some ambiguities during the interpretation of the outcomes.

# Chapter 5

# Taxonomy analysis

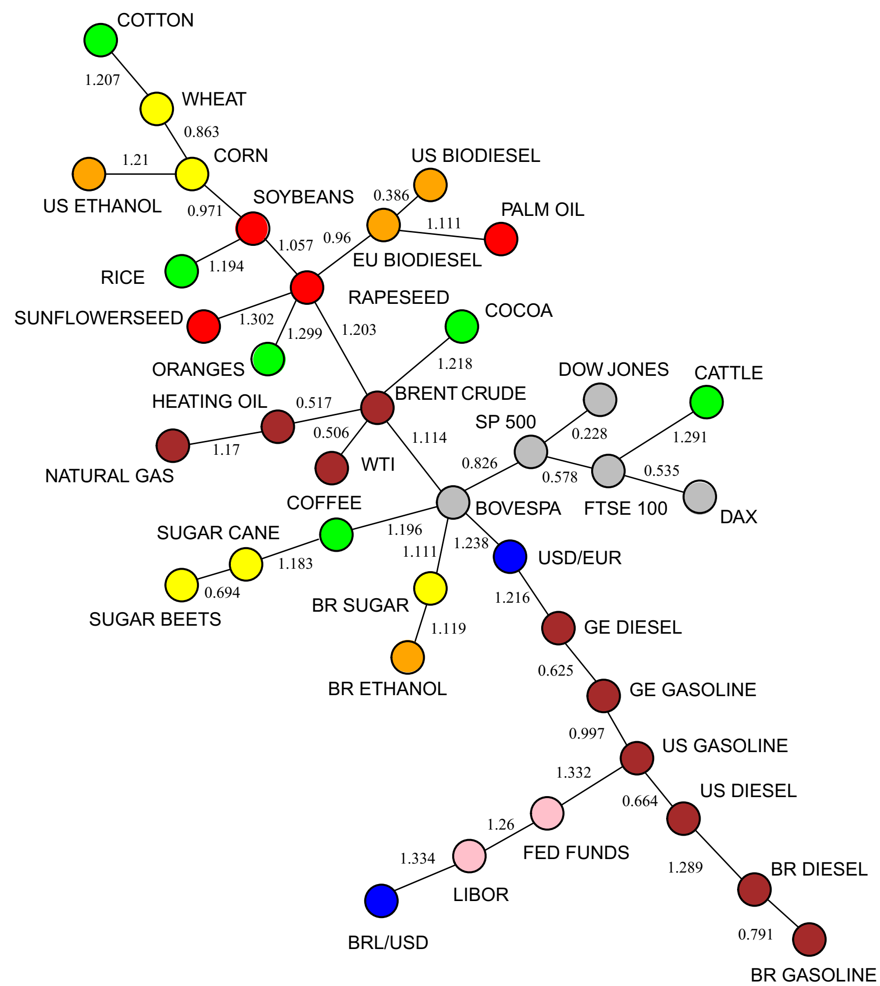
In the previous chapters, we explained the methodology of empirical analysis and described data that are going to be examined. It was also checked whether the data are well-behaved and fulfill the prerequisites and assumptions that proposed methodology requires. As long as all time series are stationary and non-normality is the issue we can deal with, nothing prevents us to approach to the taxonomy analysis itself. For every single given period, MST and HT will be conducted. All computations and graphics were made in software *R* using the package *igraph*. The periods differ in the price levels of the commodities and assets. The first reason for that is the existence of the two world food price crisis and second reason is the financial crisis followed after the collapse of Lehman Brothers. The ambition of the thesis is to discover these differences and try to interpret them. The taxonomy analysis follows work of Krištoufek et al. (2012) who brought the idea of similar patter to biofuel and commodity perspective and Filip et al. (2016) who broadened the examined dataset and structured the individual periods with the respect to the situation on the commodity market. Our aim is to undertake the analysis with a different dataset that was gathered solely for the purpose of this thesis, and to prolong the data up to February 2019 in order to receive even broader notion about the behaviour of the biofuels system over the time. First of all, we begin with the analysis of the whole time span of the dataset and consequently we examine the individual periods.

## 5.1 The entire period: 2003-2019

To achieve clear and valuable graphics, a specific colour is assigned to each commodity or asset group described in Chapter 4. Biofuels are represented by orange, fossil fuels by brown, food by green, interest rates by pink, stock indices by grey and exchange rates by blue. Biofuels feedstocks are further split to ethanol feedstock represented by yellow and biodiesel feedstock represented by red. Each edge has a numerical label that denotes the distance $d_{ij}$ among two vertices. Because of methodological properties, no correlation is negative. The distances vary in the range from 0 to $\sqrt{2}$ and the smaller is the value of the edge, the stronger is the correlation between two vertices. During the interpretation, it may happen that one may tend to overestimate a seemingly strong relationship between two neighbour vertices because of a close position in a MST graph. One should not forget that as the distance between two vertices nears to the value of 1.4, what is the approximate value of the $\sqrt{2}$, the linear relationship among them is disappearing and their connection should not be overrated. In the next paragraph, we would like to demonstrate how exactly is MST for the period 2003-2019 made according to *the Kruskal algorithm*.

The very first component of the system is the pair of stock indices, Dow Jones and SP 500 which is not surprising as they both come from the US stock market. Generally, all stock indices are closely connected and they create a cluster in every period. This particular connection is made in the distance equalled to 0.228. The second link is created at the level of 0.386 where EU biodiesel and US biodiesel are connected. Together they create a strong pair as they are related to each other in every single observed period. In the next step, another two close commodities are linked together - Brent crude oil and WTI. Wlazlowski et al. (2011) noted that WTI and Brent crude oil are the most important global price setters of crude oil and therefore it is something we might have expected. Right now, our system consists of 3 separate pairs. That is changed with the next commodity which is heating oil. It makes a link to Brent crude oil at the distance level of 0.517 which creates a triple of two crude oils (Brent and WTI) and heating oil.

**Figure 5.1: Minimum Spanning Tree − Entire time span**



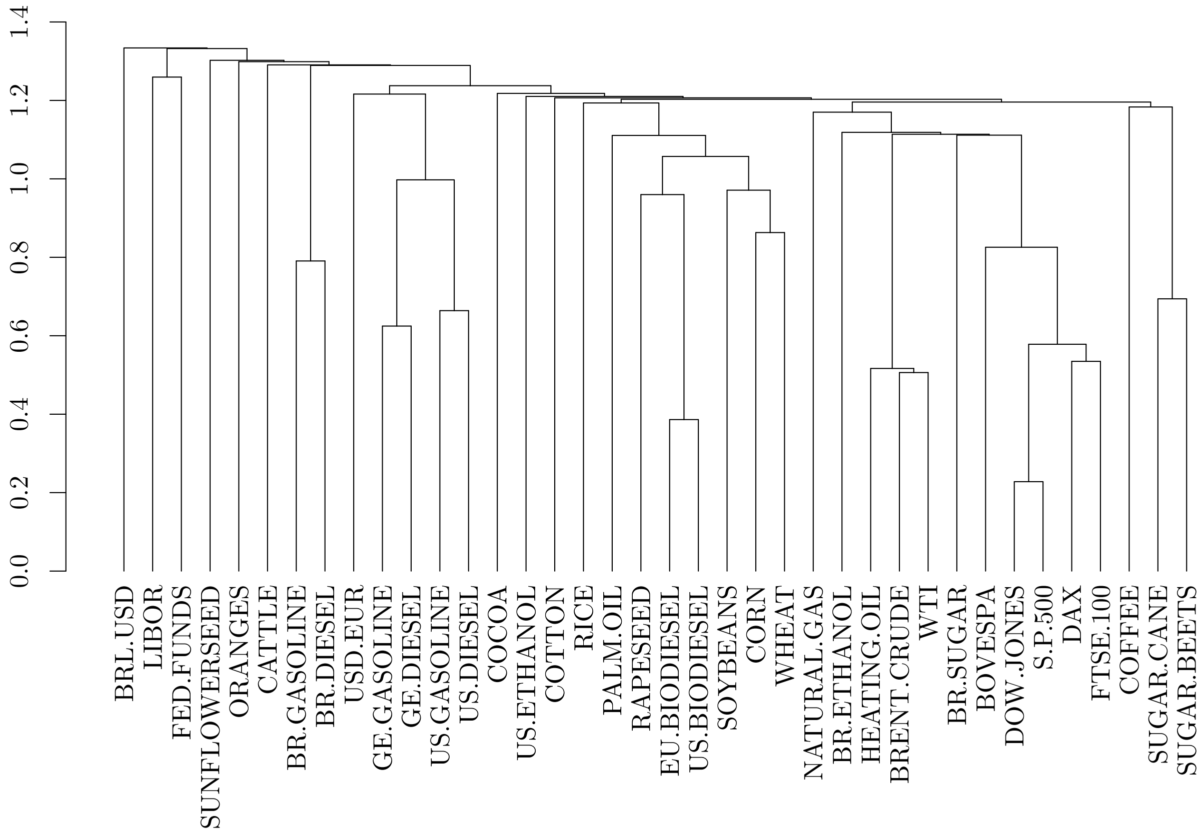

Source: Author's computation

The fourth pair of the network is established by FTSE 100 and DAX. Again, both stock indices come from Europe, so they have logically a close relationship. In the next step, it may be inspected what happens when the algorithm comes to the case of edge that connects vertices that both are already in the network. Such a new connection would establish a loop, which is strictly forbidden in MST and therefore FTSE 100 and SP 500 will not be connected and the process will continue further until the system will consist of 38 nodes and 37 connections. This system is plotted in Figure 5.1. All links that appeared in the system may be found in Appendix C, Table 5.

What looks apparent from the first look into Figure 5.1 is the clusterization of examined commodities and assets, which is the phenomenon described by Sieczka and Hołyst (2009). The clusterization means that commodities that belong to similar sector appear in a MST graph close to each other. When we put together information gained from MST in Figure 5.1 and HT in Figure 5.2, we may recognize 5 clusters. The first cluster is created by biofuels - US ethanol, EU biodiesel, US biodiesel - and their feedstocks. The second cluster consists of conventional fossil fuels (gasolines and diesels). The core of the whole network is made by the last three clusters. First, it is a cluster of stock indices which are relatively tightly connected to the cluster of oils and natural gas through the connection Bovespa - Brent crude. The last cluster of sugar crops and coffee is connected to this core also via Bovespa which plays an important role in this MST as it also connects Brazilian sugar and ethanol to the system. Hence, the Brazilian commodities are located close to each other in the MST for the period that covers the whole time span of the dataset.

All observed biofuels create a reasonable relationship with their feedstock. The strength of these relationships corresponds with the mean value of that links in the four individual periods. EU biodiesel - rapeseed are linked in the entire period at a distance of 0.96, while the mean for the four sub-periods is 0.969. The link between BR ethanol and BR sugar has in the entire period value of 1.119, while the mean for the four sub-periods is 1.108. The case of US ethanol and corn is not so straightforward, because, in the first sub-period, US ethanol and corn are not linked together. However, the compared values are also similar (1.21 for the entire period and 1.153 for the mean value of the sub-periods).

**Figure 5.2: Hierarchical Tree - The entire period**




Source: Author's computation

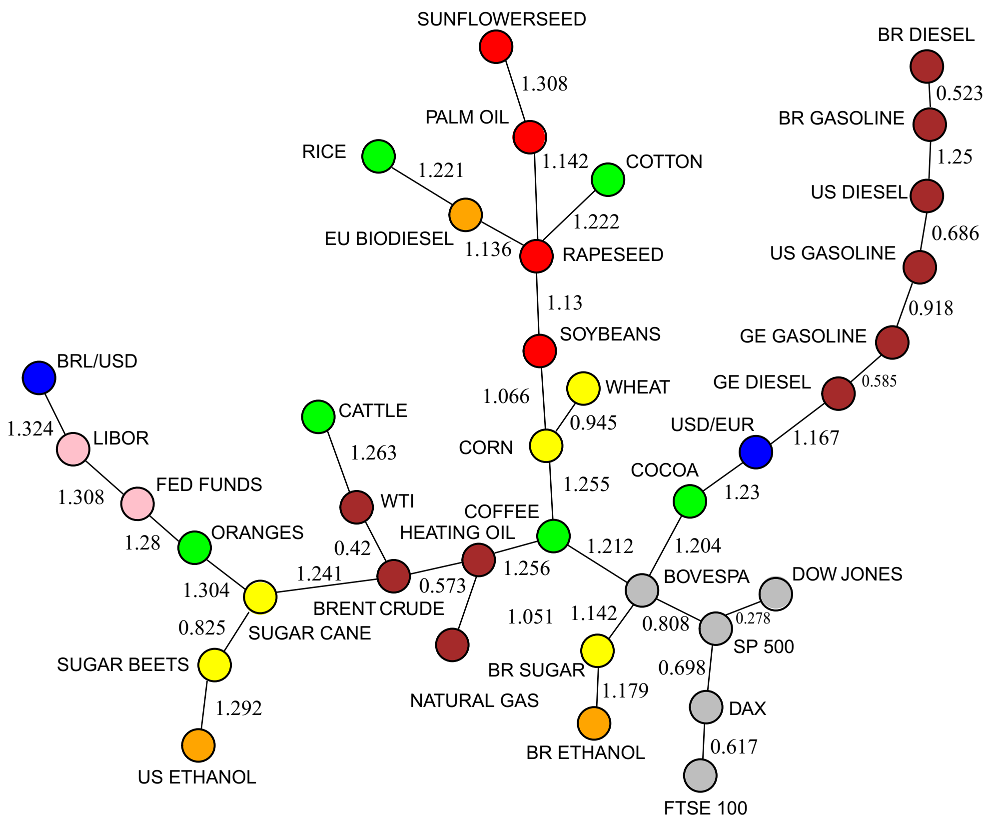
As it was stated in the methodological part, the construction of Hierarchical Tree is done in a similar manner as the construction of MST. The algorithm starts with the pair of nodes that have the lowest distance - stock indices and EU, US biodiesels. The process of choosing the lowest distances between the nodes continues until the situation known from the MST construction - 3 separate pairs of nodes: EU - US biodiesel, Dow Jones - SP 500, Brent crude - WTI. The fourth edge connects heating oil, a new member of the tree and Brent crude which has already its position in the tree. The value of the edge is 0.517 and at that distance also other commodities already related to Brent crude are connected to heating oil - at this case WTI. Every new nod or group of nods, that will be connected to some of these three commodities will be connected also to the rest of this group. In this particular example, the first commodity that joins this branch of the tree is Bovespa at the level of 1.114. This connection leads to Brent crude, although it will connect the whole branch as it was explained. As we continue adding new vertices to the system in the same fashion, we finally end up with Hierarchical Tree plotted in Figure 5.2 that works in a bottom-up manner.

## 5.2 Period I: 2003-2007

The first period is characterized by calm and relatively low food prices. The average value of FFPI index was equalled to 119.4.

The first pair of the tree is formed by EU biodiesel and US biodiesel, the distance is approaching to null which would mean that these two time series are actually the same. However, in the next periods, they will become linearly more independent and that is the reason why we keep them both in the dataset. A pair with the second lowest value 0.278 is created by two stock indices Dow Jones and SP 500. The strong relationship is showed also by crude oils, Brent and WTI, which distance equalls to 0.420. The next link brings to the network two conventional fossil fuels: Brazilian gasoline and diesel. They are connected at the distance of 0.522. The fifth connection that is added into the system is Brent crude and heating oil (0.560). Because Brent crude is already in the system, a triple WTI - Brent Crude - heating oil is established. Following two picks create two separate pairs of German diesel and gasoline (0.583) and DAX and FTSE 100 (0.611). As a result, the tree now consists of five separate pairs and one triple. Next connection among WTI and heating oil would create a loop, so we do not include the connection into the system. Proceeding in the similar manner, we arrive to the MST plotted in Figure 5.1 that consists of 38 nodes and 37 connections. The variable US biodiesel is not plotted in the graph due to almost perfect correlation with EU biodiesel which makes the graph unclear. All links that appeared in the system may be found in Appendix C, Table 6.

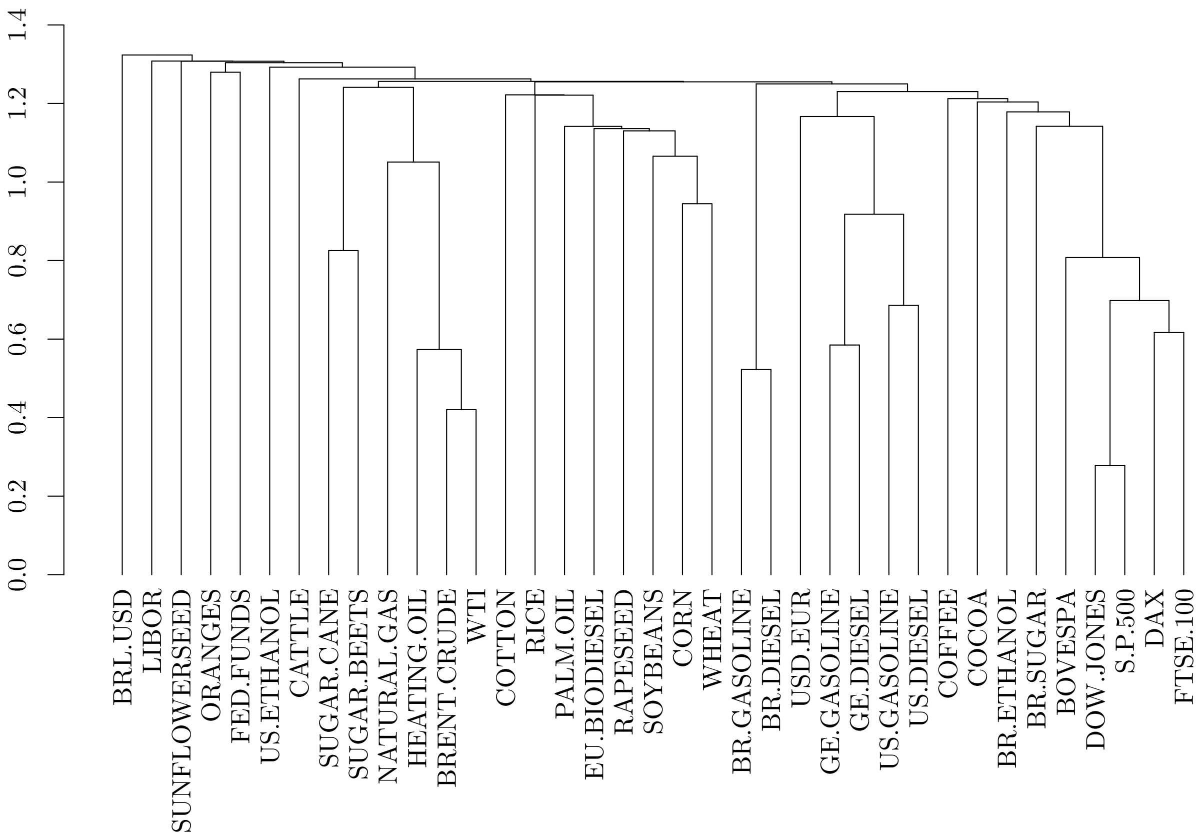
## Figure 5.3: Minimum Spanning Tree − Period I




Source: Author's computation

In the case of Figure 5.3, four clusters are created - the cluster of financial assets including stock indices and interest rates, the cluster of fossil fuels, the cluster of crude oils, which includes heating oil (a derivative of crude oil), and natural gas, direct substitute of the crude oil. The last one is the cluster of biofuels feedstocks which includes both ethanol and biodiesel crops. These four branches are connected together by coffee, nevertheless, its distances to the clusters are rather long (more than 1.2) which does not indicate a strong relationship. A bit startling is the behaviour of biofuels. Filip et al. (2016) suggest that with the low price levels (e.g. in pre-crisis period), US ethanol and biodiesel are more related to their fossil fuels and with an increase in prices, they migrate to their feedstocks. In contrast, in this case, both commodities are related to the feedstock instantly since the first period. US ethanol connection to the system is rather weak, moreover, it is related to sugar beets, which are not its main crop, but the biodiesel connection to rapeseed is direct (1.136, which rather weak link, though). This difference may be caused by using different time series of biodiesel. BR ethanol behaves as it was expected. Due to the special situation on the Brazilian fuels market, where the state-owned enterprise Petrobras has the monopolist position with no other competitors, BR ethanol is always part of the Brazilian branch of commodities. In the first period, BR ethanol is connected with BR sugar and they are linked to the system via Bovespa. The Brazilian oil market and its only market player Petrobras is well described by Silvestre et al. (2018). The average distance in MST for the first period is 1.029 which corresponds to the average correlation of 0.471,. That is the lowest value from all observed periods.

**Figure 5.4: Hierarchical tree - Period I**




Source: Author's computation

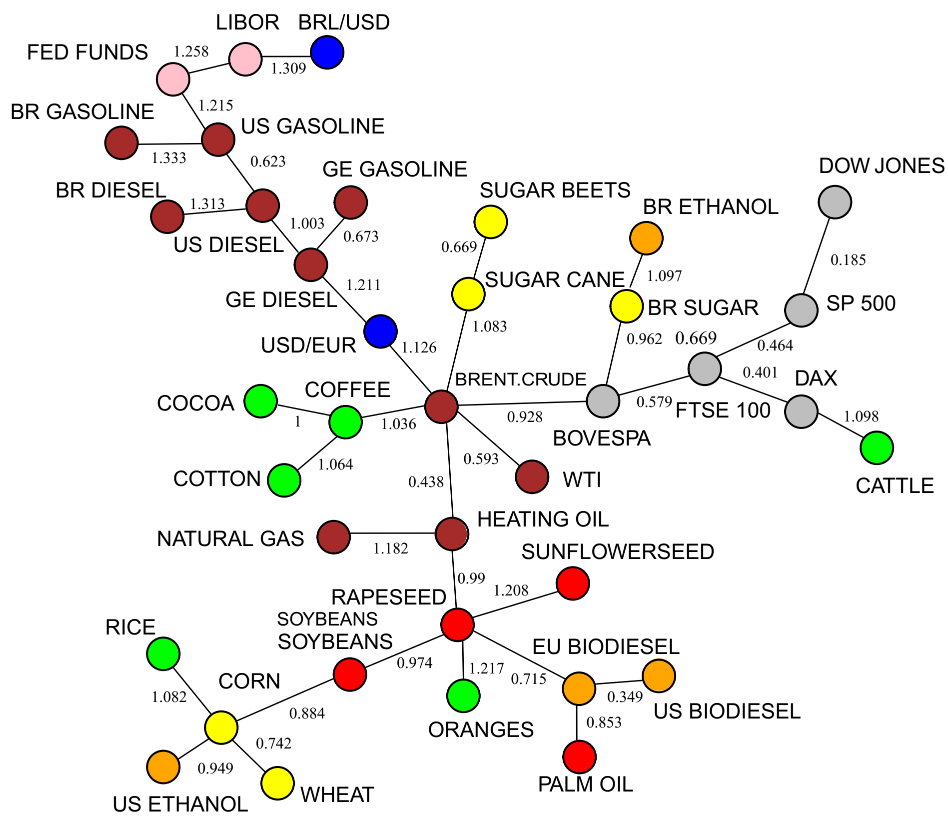
The construction of HT is described step by step for the entire period and because the sequence of the links does not differ to MST, it will not be commented again. In HT (Figure 5.4), we can observe the same situation as in MST- three relatively separated clusters of financial instruments, fossil fuels, and biofuels feedstock. The remoteness of biofuels and crude oil from the rest of the network is noticeable as these commodities join one of the three clusters rather at the end of the tree. This could mean that in this period, biofuels do not interact at the higher (weekly) frequency. Similarly to MST, US biodiesel is not plotted in the tree.

## 5.3 Period II: 2008-2011

The second period is bordered by two food prices spikes in the years 2008 and 2011. That is reflected also in the average value of FFPI. While the value of the index in the first period was 119.4, in the second period it increased to 192.0. According to Krištoufek et al. (2012), different price levels of the commodities lead to different links in the network. The weights of the links may be also affected. Such behaviour of the network is also visible from the average distance of the links of the MST. In the first period, the average distance of 1.029 was observed, whereas in the second period the distance lowered to 0.914, which corresponds to correlation of 0.58. The structure of the system changed as well.

This time, the pair with the lowest distance, equalled to 0.185, is Dow Jones and SP 500. The second pair of the tree, EU and US biodiesel, joins the system at the distance value of 0.349. The mutual independence of the biodiesels increased significantly and this process will continue in later periods. It might mean that as the market with biodiesel expands, the individual markets get more independent to each other. The next link consists of DAX and FTSE 100 and enters the system at the distance value of 0.400. Fourth link connects Brend crude and heating oil at level of 0.438. In this setup, the system is comprised by four separate pairs. This is changed with the fifth link as pairs Dow Jones - SP 500 and DAX - FTSE 100 are connected through FTSE 100 and SP 500 and create quadruple of stock indices. The next link, between DAX and SP 500, would create a loop and that is the reason why it is not included in the tree. As we would continue in the description of the algorithm, we would end up with the system graphed in Figure 5.5. All links that appeared in the system may be found in Appendix C, Table 7.

## Figure 5.5: Minimum Spanning Tree − Period II

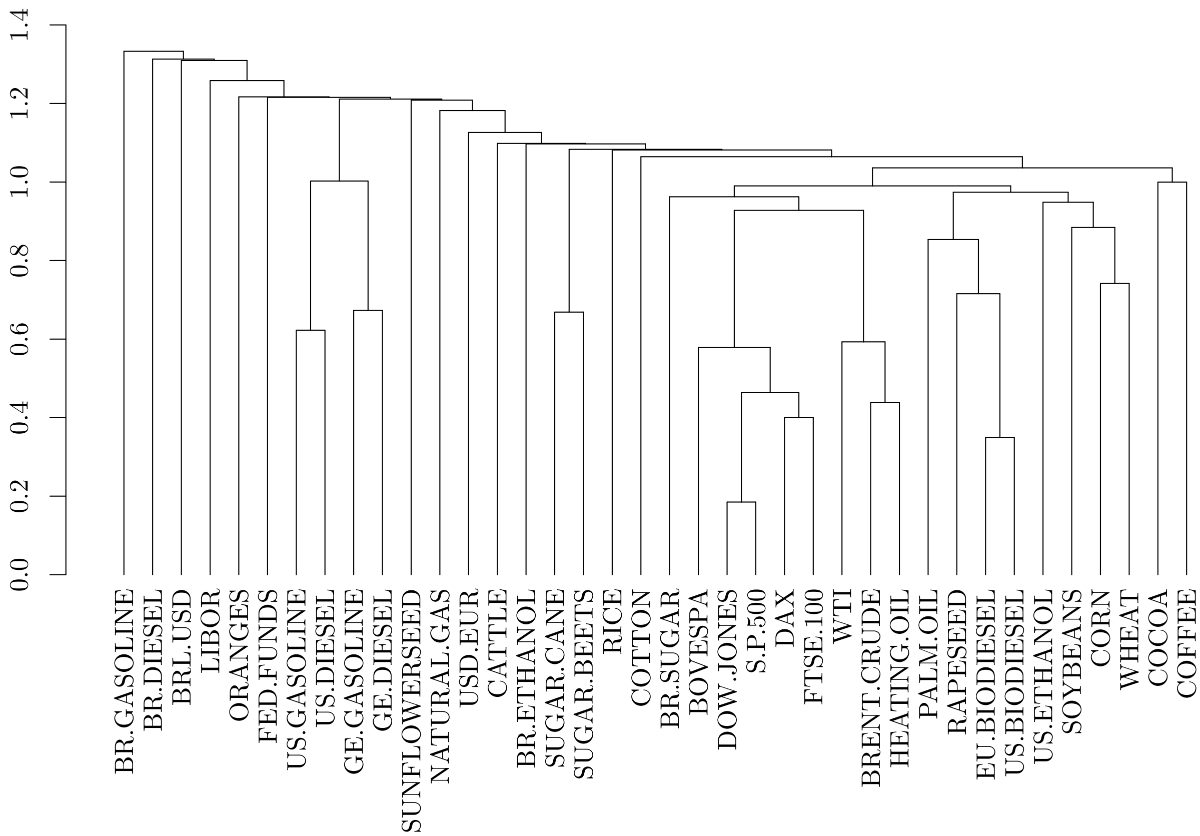



Source: Author's computation

Figure 5.5 represents MST where some notable changes to the first period may be observed. In the second period, Brent crude is the central node which connects the most branches together. This role is more logical for Brent crude than for coffee as the price of Brent crude influences prices of a large number of other commodities as suggested by Natanelov et al. (2011). Moreover, the distances which connect Brent crude to the branches decreased as compared to coffee in the first period. It is not surprising that the Brent crude is especially close to WTI and heating oil. Another important change is the position of US ethanol which migrated next to its direct feedstock - corn. After this move, every single biofuel is directly connected to its organic feedstock (BR ethanol is tapped to BR Sugar, not sugar cane but still we may consider it as a feedstock). The strongest is the link of the pair EU and US biodiesel to rapeseed (0.715) and palm oil (0.853), followed by US ehanol to corn (0.949), and BR ethanol to BR Sugar (1.097). Soybeans remain in the role of the connector of biodiesel feedstock and US ethanol feedstock which together form a cluster. A cluster is also formed by stock indices. This time, conventional fossil fuels are closer to group of crude oils, connected with USD/EUR exchange rate. An interesting group is created by triple coffee, cocoa, and cotton, which are tied both closer together and closer to the rest of the system than in the previous period. This is probably caused by higher price levels of the food. Interest rates are at this period connected to fossil fuels, but, again, with the only mild link.

When we look at Hierarchical Tree for the second period in Figure 5.6, we may notice the strong links in the middle of the tree. Groups of stock indices and crude oils with heating oil connect together at level 0.927. This cluster creates the core of the tree. Other components of the system join the core later in the process. In the perspective of biofuels, biodiesel group and ethanol group join each other at level of 0.884 and they are immediately linked to the core of the system. The most remote is BR ethanol with the connection at 1.097.

**Figure 5.6: Hierarchical tree - Period II**



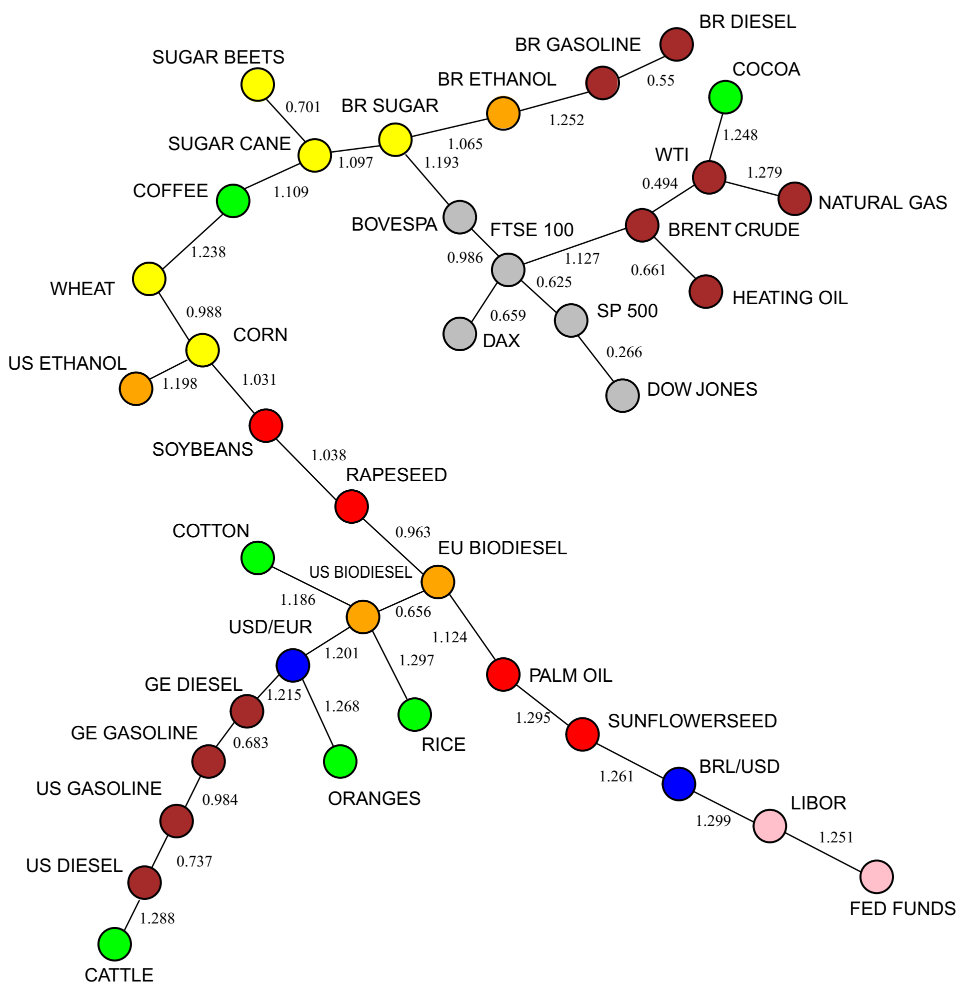

Source: Author's computation

## 5.4 Period III: 2011-2015

After two significant price hikes in the second period, food prices remained high after the second price spike in 2011. However, prices became less volatile and gradually decreased. These conditions created once again a new environment for the development of biofuels and related commodities. The average distance in MST graph for the third period equalled to 1.014 and therefore lower overall interconnection among the members of the system may be expected. MST graph is created in the same manner as in the two previous periods, so the exact procedure will not be described. All links that appeared in the system may be found in Appendix C, Table 8.

Figure 5.7 delivers MST graph of third period in which one crucial feature

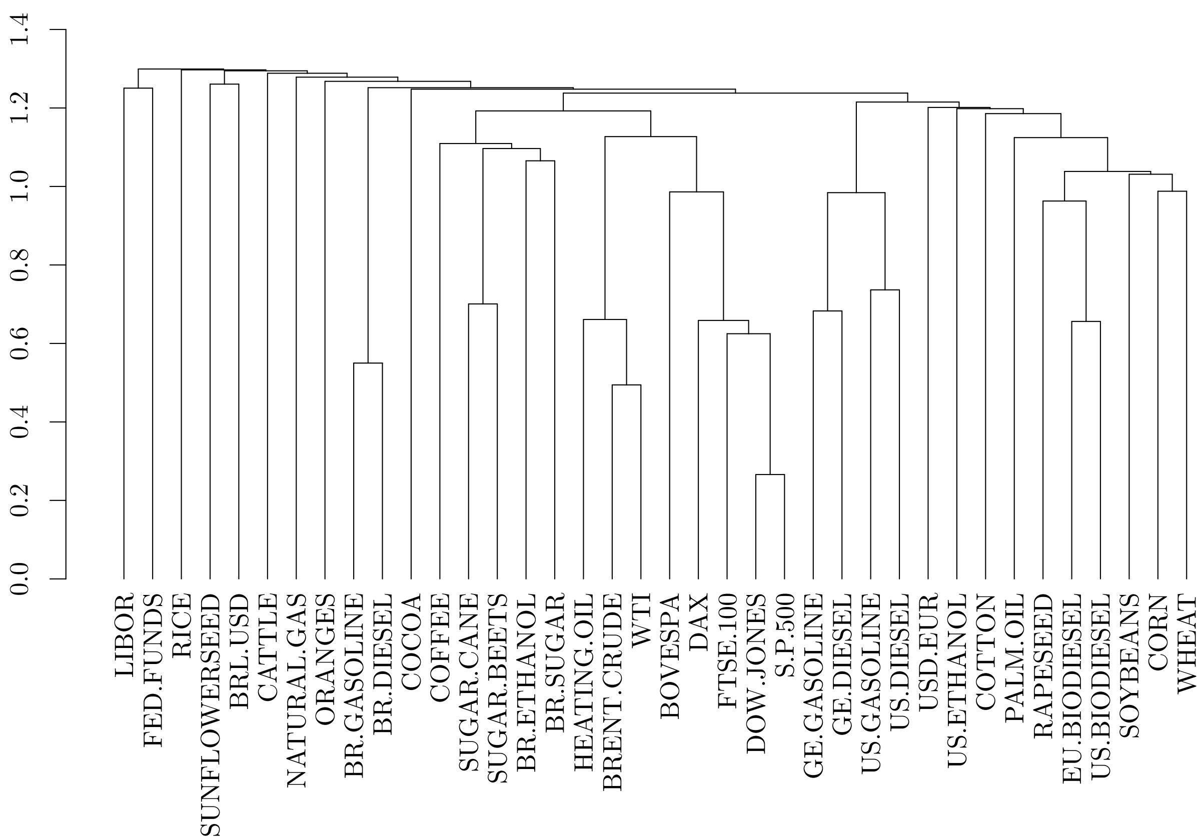
## Figure 5.7: Minimum Spanning Tree − Period III




Source: Author's computation

is preserved. Every single biofuel is directly linked to its feedstock, in spite of the lower strength than in the previous period. This is important mainly for US ethanol which in the first period did not form any relationship with corn or wheat. On the other hand, when the tree is compared to the previous period′s one, the branches are not tied together with a central nod which was Brent crude.

**Figure 5.8: Hierarchical tree - Period III**



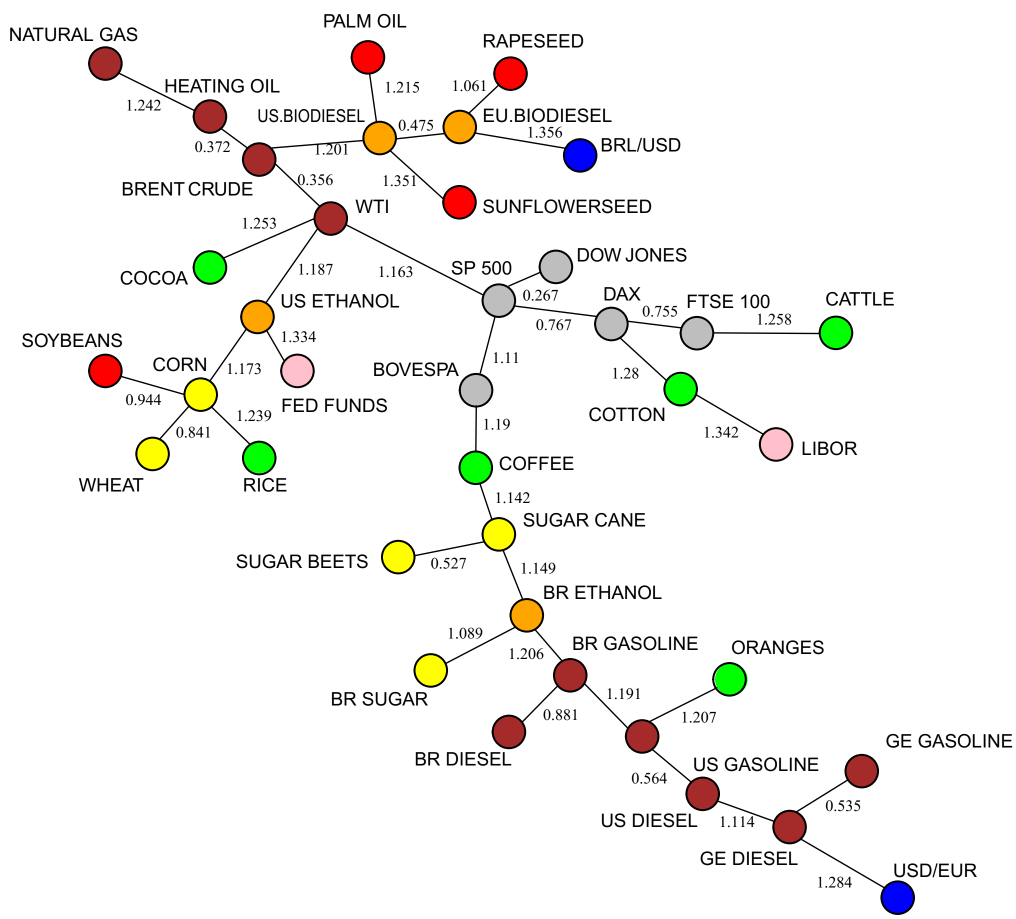

Source: Author's computation

The tree is separated into four branches as visible from HT tree in Figure 5.8. The first branch consists of US ethanol, biodiesels, and their related feedstocks. This is something we might have firstly observed in the previous period in which the prices of food increased sharply and the cluster still appears in the third period. Similarly, as in the second period, the strongest connection in the biofuel-crop relationship was recognized by EU biodiesel and rapeseed 0.965 which is a decrease compared to the second period. The same trend applies to other biofuels. Moreover, this cluster works as a connector to other parts of the network. On the right part of the system, Brazilian commodities such as sugars, coffee, and BR ethanol lay together with bigger cluster of stock indices and crude oils which are connected via Brend crude - FTSE 100 at the level of 1.126. The fourth cluster is created from gasolines and diesels supplied in all three markets and it is located in the very left of MST figure 5.7. It is noteworthy that in the second and third period, two branches - crude oils and stock indices - form together a tighter cluster than in the first and fourth period. This corresponds with the development of oil prices. In the second and third period, prices of crude oil were significantly higher than in the first and last period. While the average WTI price in period two and three was 83.8 US dollars per gallon, the average price in period one and four was 55.8 US dollars per gallon. A mini, green coloured, cluster of food, which has appeared in the previous period, disappeared. Single food commodities may be found all over the graph. As this is also true for the period one and four, the period two, with volatile food prices, forms an exception in the behaviour of the food group.

## 5.5 Period IV: 2016-2019

The fourth period is characterized by stable food prices that do not experience any significant fluctuations. The average length of the links in the system rose, in comparison with the third period ,to the value of 1.015 which nears to the value of the first period that equalled to 1.018. This value confirms the hypothesis that, in the case of lower prices of the food according to FAO index, prices in the biofuels-fuels-food system get more independent to each other.

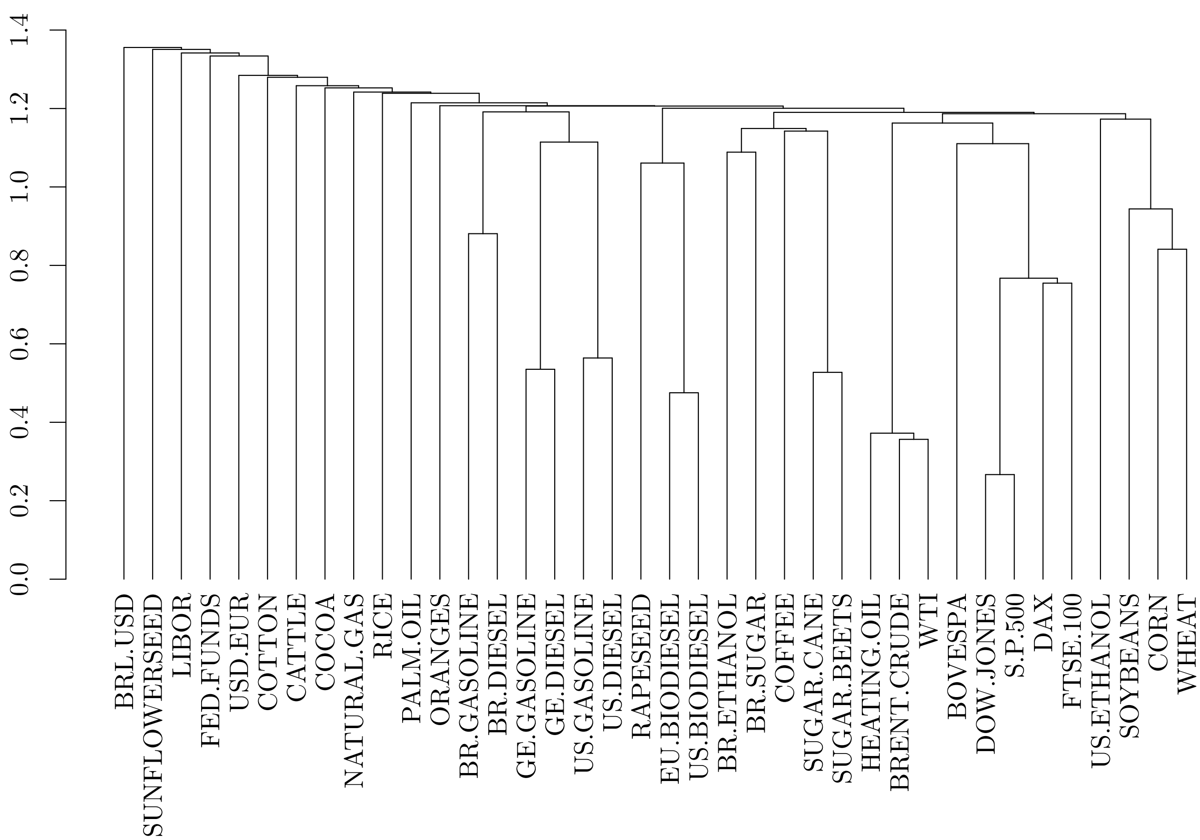
## Figure 5.9: Minimum Spanning Tree − Period IV




Source: Author's computation

On the contrary, as prices of food rise, individual members of the system influence others with a higher strength. Therefore, the network is more prone to an endogenous price shock that affects members of the network in time when prices are at a higher level. Already the first look into MST in Figure 5.9 reveals that it is less clear to recognize particular clusters in the tree. That is confirmed by HT in Figure 5.10. Clusters are created at higher levels of distance, often around the value 1.2 which is a slightly higher level than in the previous period. If we look more precisely, we can count up to six individual clusters, namely: gasolines and diesels, Brazilian ethanol and its crops, US ethanol and its crops together with soybeans, stock indices, crude oils, and biodiesels and their crops. The biggest difference of the fourth period to the periods of the higher prices is the disappearance of the group vegetable oils that were traditionally linked close to the biodiesel. In this period, only rapeseed and palm oil are directly connected to biofuels. Soybeans are now linked quite strongly to corn (0.943). Sunflower seeds are weakly connected to crude oil branch. Biofuels behave in a similar manner as in the previous period. On the one hand, the link in the pair EU biodiesel - rapeseed became significantly weaker, by 0.148, on the other hand, links between US ethanol - corn and BR ethanol and BR sugar strengthen slightly (0.015 and 0.031). The tightness of US biodiesel to its European variant increased by 0.169. From this, it might be concluded that the price level decrease from third to the fourth period has the more significant impact on biodiesel than on ethanol. All links that appeared in the system may be found in Appendix C, Table 9.

**Figure 5.10: Hierarchical tree - Period IV**




Source: Author's computation

# Conclusion

The main purpose of the thesis was to assess the price relationships in the food-fuels-biofuels system. In order to bring some insight into this broad topic, the appropriate methodology was utilized. The 38 price time series of various commodities and assets that are related to this system were analyzed with Minimum Spanning Tree and Hierarchical Tree methods. The analysis was conducted for various time periods which reflected the development of the world food prices. And as the results revealed, the food prices are critical for the behaviour of relationships in the system. It appears that there exists a link between the level of food prices and the degree of price interconnection among commodities and assets. In the periods of higher food prices, the correlation among members of the system is higher than in the periods of lower food prices. This trend can be measured on the average distance in MST. The first period is characterized by the lowest food prices[1] from all observed periods and the average distance of MST reached a value of 1.029. The second period brought a sharp increase in food prices[2] and two food prices hikes. That is reflected in the average distance of MST which was equalled to 0.914. The last period is defined by the shift back to lower food prices[3] and the average distance of MST once again reacted and increased to 1.017. All stated values are placed on the scale from 0 to $\sqrt{2}$ where 0 means perfect correlation and $\sqrt{2}$ represents no correlation at all.

The described trend can be applied also to the main objective of the thesis - biofuels - and their relationship to feedstock. Despite the fact that

[1] FFPI = 119.3
[2] FFPI = 192.0
[3] FFPI = 168.0

we cannot comment on the direction of the relationship, we might add some arguments into the discussion whether the increasing biofuels production drives up the prices of food from a distance metrics point of view. In the first period, links of biofuels to their feedstock are rather weak. Biodiesel and Brazilian ethanol reached the maximum distance[4] to feedstock, US ethanol is not linked to its main feedstock at all. The second period with its food price spikes brings a sharp increase in the strength of the links, as the relationships between biodiesel-rapeseed and US ethanol-corn were the strongest[5] from all studied periods. If this trend had continued also in the next periods, in which biofuels production soared, it would have been a clear signal for supporting the theory that biofuels cause higher prices of food. However, this did not happen. With the exception of Brazilian ethanol, which is as we described specific for the fuel policy in Brazil, biofuel-feedstock relationships became again weaker in the third[6] and fourth[7] period, far from the all time maximum strength in the second period. Therefore, the thesis results suggest that there is not a clear pattern in the biofuels-food relationship with respect to the size of biofuels production. Such a result is consistent with the recent research that claim that, despite some role of biofuels, the increases in oil prices, changes in exchange rates, trade policies or speculations in food commodities are the main contributors in the increase of agricultural prices since 2004 (Popp et al. (2018), Kline et al. (2017)).

Another contribution of the thesis is a comprehensive overview of the behaviour of biofuels-related commodities for more than 17 years. We can confirm the results of Filip et al. (2016) regarding US and Brazilian ethanol. Our analysis suggests the similar attitude of these biofuels toward their feedstock. By contrast, biodiesel behaved in a different manner as it was quite strongly related to rapeseed already the first period. Moreover, except for the first period, biodesel has the strongest link to its feedstock among all studied biofuels.

[4] biodiesel-rapeseed: 1.136, BR ethanol-sugar: 1.179

[5] bidiesel-rapessed: 0.715, US ethanol-corn: 0.949

[6] biodiesel-rapeseed: 0.963, US ethanol-corn: 1.198

[7] biodiesel-rapeseed: 1.061, US ethanol-corn: 1.173

The thesis discusses methods of MST and HT which are used for the advantage of their straightforward results. However, thanks to its relative simplicity, it has certain limitations. The most importantly, we cannot study the direction between the discovered links. Causality is in the thesis neglected as well. Many agricultural commodities are heavily dependent on crude oil prices because significant amount of oil is used in their production. The detailed role of crude oil in the relationships of biofuels and their feedstock could be an topic for additional research in this area.

On the other hand, the thesis provides the general overview about the world of biofuels, shows the most important relationships and puts them into the perspective of the world food prices. Moreover, thesis brings well-organized description of the dataset which is composed from various sources. The dataset is unique in the number of biofuels related commodities and also in the length of time span that is covered. This dataset could be a good starting point for further time series analysis in the area of biofuels.

# Bibliography


Abdelradi, F. and Serra, T. (2015), ‘Food–energy nexus in europe: Price volatility approach’, *Energy Economics* **48**, 157–167.

*A Brief History of Octane in Gasoline: From Lead to Ethanol* (2016). **URL:** *https://www.eesi.org/papers/view/fact-sheet-a-brief-history-of-octane*

Asuero, A., Sayago, A. and Gonzalez, A. (2006), ‘The correlation coefficient: An overview’, *Critical reviews in analytical chemistry* **36**(1), 41–59.

Banschbach, V. S. and Letovsky, R. (2010), ‘The use of corn versus sugarcane to produce ethanol fuel: a fermentation experiment for environmental studies’, *The American Biology Teacher* **72**(1), 31–36.

Bentivoglio, D., Finco, A., Bacchi, M. R. P., Spedicato, G. et al. (2014), ‘European biodiesel market and rapeseed oil: what impact on agricultural food prices’, *International Journal of Global Energy Issues* **37**(5-6), 220–235.

Bonanno, G., Caldarelli, G., Lillo, F., Micciche, S., Vandewalle, N. and Mantegna, R. N. (2004), ‘Networks of equities in financial markets’, *The European Physical Journal B* **38**(2), 363–371.

Bracmort, K. (2018), ‘The renewable fuel standard (rfs): An overview’, *Congressional Research Service, Washington, DC, US* .

Busse, S., Brümmer, B. and Ihle, R. (2010*a*), Investigating rapeseed price volatilities in the course of the food crisis, Technical report.

Busse, S., Brümmer, B. and Ihle, R. (2010*b*), The pattern of integration between fossil fuel and vegetable oil markets: The case of biodiesel in germany, Technical report.

Busse, S., Brümmer, B. and Ihle, R. (2012), 'Price formation in the german biodiesel supply chain: a markov-switching vector error-correction modeling approach', *Agricultural Economics* **43**(5), 545–560.

Capitani, D. H. D., Tonin, J. M. and Cruz, J. C. (2017), 'Integration and hedging efficiency between the Brazilian and the US ethanol markets'.

De Gorter, H., Drabik, D., Kliauga, E. M. and Timilsina, G. R. (2013), *An economic model of Brazil's ethanol-sugar markets and impacts of fuel policies*, The World Bank.

De Gorter, H., Drabik, D. and Timilsina, G. R. (2013), *The effect of biodiesel policies on world oilseed markets and developing countries*, The World Bank.

Dickey, D. A. and Fuller, W. A. (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American statistical association* **74**(366a), 427–431.

Diestel, R. (2012), 'Graph theory, volume 173 of', *Graduate texts in mathematics* p. 7.

Drabik, D., Ciaian, P. and Pokrivčák, J. (2016), 'The effect of ethanol policies on the vertical price transmission in corn and food markets', *Energy Economics* **55**, 189–199.

Du, X. and Hayes, D. J. (2012), 'The impact of ethanol production on us and regional gasoline markets: an update to 2012'.

Dutta, A. (2018), 'Cointegration and nonlinear causality among ethanol-related prices: evidence from Brazil', *GCB Bioenergy* **10**(5), 335–342.

EIA (2019), 'U.s. energy information administration - eia - independent statistics and analysis', *Monthly Biodiesel Production Report - Energy Inform-*

*ation Administration* .

**URL:** *https://www.eia.gov/biofuels/biodiesel/production/*

FAO (2018), 'Oecd-fao agricultural outlook 2018-2027'.

**URL:** *http://www.fao.org/publications/oecd-fao-agricultural-outlook/2018-2027/en/*

FAS (2018*a*), 'Brazil biofuels annual', *Annual Report 2018* .

FAS (2018*b*), 'EU-28: Biofuels annual', *EU-28: Biofuels Annual — USDA Foreign Agricultural Service* .

FAS (2018*c*), 'EU-28: Biofuels mandates', *Biofuel Mandates in the EU by Member State in 2018* .

Fernandez-Perez, A., Frijns, B. and Tourani-Rad, A. (2016), 'Contemporaneous interactions among fuel, biofuel and agricultural commodities', *Energy Economics* **58**, 1–10.

Filip, O., Janda, K., Kristoufek, L. and Zilberman, D. (2016), 'Dynamics and evolution of the role of biofuels in global commodity and financial markets', *Nature Energy* **1**(12), 16169.

Filip, O., Janda, K., Kristoufek, L. and Zilberman, D. (2017), 'Food versus fuel: An updated and expanded evidence', *Energy Economics* .

Green, T. R., Kipka, H., David, O. and McMaster, G. S. (2018), 'Where is the usa corn belt, and how is it changing?', *Science of the Total Environment* **618**, 1613–1618.

Harris, R. I. (1992), 'Testing for unit roots using the augmented dickey-fuller test: Some issues relating to the size, power and the lag structure of the test', *Economics letters* **38**(4), 381–386.

Hassouneh, I., Serra, T., Goodwin, B. K. and Gil, J. M. (2012), 'Non-parametric and parametric modeling of biodiesel, sunflower oil, and crude oil price relationships', *Energy Economics* **34**(5), 1507–1513.

Havlicek, L. L. and Peterson, N. L. (1976), ‘Robustness of the pearson correlation against violations of assumptions’, *Perceptual and Motor Skills* **43**(3_suppl), 1319–1334.

Janda, K. and Krištoufek, L. (2019), ‘The relationship between fuel and food prices: Methods, outcomes, and lessons for commodity price risk management’.

Jarque, C. M. and Bera, A. K. (1980), ‘Efficient tests for normality, homoscedasticity and serial independence of regression residuals’, *Economics letters* **6**(3), 255–259.

Khanna, M., Nuñez, H. M. and Zilberman, D. (2016), ‘Who pays and who gains from fuel policies in Brazil?’, *Energy Economics* **54**, 133–143.

Kline, K. L., Msangi, S., Dale, V. H., Woods, J., Souza, G. M., Osseweijer, P., Clancy, J. S., Hilbert, J. A., Johnson, F. X., McDonnell, P. C. et al. (2017), ‘Reconciling food security and bioenergy: priorities for action’, *Gcb Bioenergy* **9**(3), 557–576.

Krištoufek, L., Janda, K. and Zilberman, D. (2012), ‘Correlations between biofuels and related commodities before and during the food crisis: A taxonomy perspective’, *Energy Economics* **34**(5), 1380–1391.

Krištoufek, L., Janda, K. and Zilberman, D. (2016), ‘Comovements of ethanol-related prices: evidence from Brazil and the usa’, *Gcb Bioenergy* **8**(2), 346–356.

Kruskal, J. B. (1956), ‘On the shortest spanning subtree of a graph and the traveling salesman problem’, *Proceedings of the American Mathematical society* **7**(1), 48–50.

Kwiatkowski, D., Phillips, P. C., Schmidt, P. and Shin, Y. (1992), ‘Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?’, *Journal of econometrics* **54**(1-3), 159–178.

Lautier, D. and Raynaud, F. (2012), 'Systemic risk in energy derivative markets: a graph-theory analysis', *The Energy Journal* pp. 215–239.

Mantegna, R. N. (1999), 'Hierarchical structure in financial markets', *The European Physical Journal B-Condensed Matter and Complex Systems* **11**(1), 193–197.

Myers, R. J., Johnson, S. R., Helmar, M. and Baumes, H. (2014), 'Long-run and short-run co-movements in energy prices and the prices of agricultural feedstocks for biofuel', *American Journal of Agricultural Economics* **96**(4), 991–1008.

Natanelov, V., Alam, M. J., McKenzie, A. M. and Van Huylenbroeck, G. (2011), 'Is there co-movement of agricultural commodities futures prices and crude oil?', *Energy Policy* **39**(9), 4971–4984.

Nešetřil, Matoušek, J. et al. (2009), *Invitation to discrete mathematics*, Oxford University Press.

Onnela, J.-P. (2002), 'Taxonomy of financial assets', *Unpublished master's thesis). Dep. of Electrical and Communications Engineering, Helsinki University of Technology* .

Pal, D. and Mitra, S. K. (2017), 'Time-frequency contained co-movement of crude oil and world food prices: A wavelet-based analysis', *Energy Economics* **62**, 230–239.

Perez, I. and Wire, C. (2013), 'Climate change and rising food prices heightened arab spring', *Scientific American* **4**(03), 2013.

Pimentel, D., Marklein, A., Toth, M. A., Karpoff, M. N., Paul, G. S., McCormack, R., Kyriazis, J. and Krueger, T. (2009), 'Food versus biofuels: environmental and economic costs', *Human ecology* **37**(1), 1.

Piroli, G., Rajcaniova, M., Ciaian, P. and Kancs, d. (2014), From a rise in b to a fall in c? environmental impact of biofuels, Technical report, EERI Research Paper Series.

Plevin, R. J., Jones, A. D., Torn, M. S. and Gibbs, H. K. (2010), 'Greenhouse gas emissions from biofuels™ indirect land use change are uncertain but may be much greater than previously estimated'.

Popp, J., Kot, S., Lakner, Z. and Oláh, J. (2018), 'Biofuel use: Peculiarities and implications.', *Journal of Security & Sustainability Issues* **7**(3).

Rajcaniova, M., Kancs, d. and Ciaian, P. (2014), 'Bioenergy and global land-use change', *Applied Economics* **46**(26), 3163–3179.

*Renewables Global Status Report* (2018).
**URL:** *http://www.ren21.net/status-of-renewables/global-status-report/*

Saghaian, S. H. (2010), 'The impact of the oil sector on commodity prices: Correlation or causation?', *Journal of Agricultural and Applied Economics* **42**(3), 477–485.

Searchinger, T., Heimlich, R., Houghton, R. A., Dong, F., Elobeid, A., Fabiosa, J., Tokgoz, S., Hayes, D. and Yu, T.-H. (2008), 'Use of us croplands for biofuels increases greenhouse gases through emissions from land-use change', *Science* **319**(5867), 1238–1240.

Serra, T. and Zilberman, D. (2013), 'Biofuel-related price transmission literature: A review', *Energy Economics* **37**, 141–151.

Shapiro, S. S. and Wilk, M. B. (1965), 'An analysis of variance test for normality (complete samples)', *Biometrika* **52**(3/4), 591–611.

Sieczka, P. and Hołyst, J. A. (2009), 'Correlations in commodity markets', *Physica A: Statistical Mechanics and its Applications* **388**(8), 1621–1630.

Silvestre, H. C., Gomes, R. C., Lamba, J. R. and Correia, A. M. (2018), 'Implementation of brazil's energy policy through the national oil company: From institutional chaos to strategic order', *Energy policy* **119**, 87–96.

Singh, B. P. (2013), *Biofuel crops: production, physiology and genetics*, CABI.

*Statistical Review of World Energy* (2018).

**URL:** *https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html*

Tabak, B., Serra, T. and Cajueiro, D. (2010), 'Topological properties of commodities networks', *The European Physical Journal B* **74**(2), 243–249.

Tadasse, G., Algieri, B., Kalkuhl, M. and Von Braun, J. (2016), Drivers and triggers of international food price spikes and volatility, *in* 'Food price volatility and its implications for food security and policy', Springer, Cham, pp. 59–82.

Taghizadeh-Hesary, F., Rasoulinezhad, E. and Yoshino, N. (2019), 'Energy and food security: Linkages through price volatility', *Energy Policy* **128**, 796–806.

Timilsina, G. R. and Shrestha, A. (2010), *Biofuels: markets, targets and impacts*, The World Bank.

Vacha, L. and Barunik, J. (2012), 'Co-movement of energy commodities revisited: Evidence from wavelet coherence analysis', *Energy Economics* **34**(1), 241–247.

Vacha, L., Janda, K., Kristoufek, L. and Zilberman, D. (2013), 'Time–frequency dynamics of biofuel–fuel–food system', *Energy Economics* **40**, 233–241.

Verbeek, M. (2008), *A guide to modern econometrics*, John Wiley & Sons.

Wlazlowski, S., Hagströmer, B. and Giulietti, M. (2011), 'Causality in crude oil prices', *Applied Economics* **43**(24), 3337–3347.

Yang, Y., Bae, J., Kim, J. and Suh, S. (2012), 'Replacing gasoline with corn ethanol results in significant environmental problem-shifting', *Environmental science & technology* **46**(7), 3671–3678.

Zhang, Z., Lohr, L., Escalante, C. and Wetzstein, M. (2010), 'Food versus fuel: What do prices tell us?', *Energy policy* **38**(1), 445–451.

Zilberman, D., Hochman, G., Rajagopal, D., Sexton, S. and Timilsina, G. (2012), 'The impact of biofuels on commodity food prices: Assessment of findings', *American Journal of Agricultural Economics* **95**(2), 275–281.

# List of Tables

# List of Figures

# Appendix

## Appendix A

## Data overview and Descriptive statistics

Table 1: Data overview

| NAME | TICKER | SOURCE | TYPE |
|---|---|---|---|
| EU BIODIESEL | - | Reuters | Spot, RME |
| US BIODIESEL | - | Reuters | spot prices, RME |
| US ETHANOL | ETHNNYPR | Bloomberg | Spot, FOB, anhydrous ethanol |
| BR ETHANOL | - | CEPEA | Spot, anhydrous ethnanol |
| BRENT CRUDE | LCOc1 | Reuters | 1st month futures, ICE |
| WTI | CLc1 | Reuters | 1st month future, NYMEX |
| NATURAL GAS | NGc1 | Reuters | 1st month futures, NYMEX |
| HEATING OIL | HOc1 | Reuters | 1st month futures, NYMEX |
| US GASOLINE | - | EIA | Regular gasoline retail prices |
| US DIESEL | - | EIA | Diesel retail prices |
| GE GASOLINE | - | Reuters | Unleaded retail gasoline prices |
| GE DIESEL | - | Reuters | Diesel retail prices |
| BR GASOLINE | - | ANP | Average retail prices |
| BR DIESEL | - | ANP | Average retail prices |
| CORN | Cc1 | Reuters | 1st month futures, CBOT |

| | | | |
|---|---|---|---|
| WHEAT | Wc1 | Reuters | 1st month futures, CBOT |
| SUGAR CANE | SBc1 | Reuters | 1st month futures, ICE |
| SUGAR BEETS | QW1 COMDTY | Bloomberg | 1st month futures, LIFFE |
| BR SUGAR | - | CEPEA | ESLQ Crystal sugar price index |
| PALM OIL | 1FCPOc1 | Reuters | 1st month futures, Bursa Malaysia |
| RAPESEED | COMc1 | Reuters | 1st month futures, Euronext |
| SOYBEANS | Sc1 | Reuters | 1st month futures, CBOT |
| SUNFLOWERSEED | SUFc1 | Reuters | 1st month futures, Johannesburg exchange |
| COCOA | CCc1 | Reuters | 1st month futures, ICE |
| COFFEE | KCc1 | Reuters | 1st month futures, ICE |
| RICE | RRc1 | Reuters | 1st month futures, CBOT |
| ORANGE JUICE | OJc1 | Reuters | 1st month futures, ICE |
| CATTLE | FCc1 | Reuters | Feeder cattle,1st month futures,CME |
| COTTON | CTc2 | Reuters | 1st month futures, ICE |
| BOVESPA | .BVSP | Reuters | Sao Paulo SE Bovespa Index |
| DAX | .GDAXI | Reuters | Deutsche Boerse DAX Index |
| DOW JONES | .DJI | Reuters | Dow Jones Industrial Average Index |
| FTSE 100 | .FTSE | Reuters | The Financial Times Stock Exchange 100 Index |
| SP 500 | .SPX | Reuters | Standard & Poors 500 |
| BRL/USD | - | FRED | - |
| USD/EUR | - | FRED | - |
| LIBOR | - | FRED | 3-Month London Interbank Offered Rate |
| FED FUNDS | - | FRED | Effective Federal Funds Rate |

Source: Author's computation

Table 2: Descriptive statistics

| | Mean | Min | Max | Median | SD |
|---|---|---|---|---|---|
| EU.BIODIESEL | 0.0004 | -0.1515 | 0.1135 | 0.0000 | 0.0237 |
| US.BIODIESEL | 0.0002 | -0.1515 | 0.1135 | 0.0000 | 0.0247 |
| US.ETHANOL | -0.0002 | -0.4159 | 0.2734 | 0.0000 | 0.0518 |
| BR.ETHANOL | 0.0010 | -0.2951 | 0.2206 | 0.0020 | 0.0412 |
| BRENT.CRUDE | 0.0010 | -0.2971 | 0.2010 | 0.0035 | 0.0451 |
| WTI | 0.0007 | -0.3122 | 0.2412 | 0.0039 | 0.0490 |
| NATURAL.GAS | -0.0007 | -0.2866 | 0.2462 | -0.0027 | 0.0689 |
| HEATING.OIL | 0.0011 | -0.2070 | 0.1689 | 0.0000 | 0.0439 |
| US.GASOLINE | 0.0007 | -0.0892 | 0.1613 | 0.0000 | 0.0196 |
| US.DIESEL | 0.0009 | -0.0928 | 0.1146 | 0.0000 | 0.0171 |
| GE.GASOLINE | 0.0007 | -0.2290 | 0.2438 | 0.0000 | 0.0431 |
| GE.DIESEL | 0.0009 | -0.1971 | 0.2120 | 0.0000 | 0.0384 |
| BR.GASOLINE | 0.0009 | -0.0492 | 0.0805 | 0.0000 | 0.0078 |
| BR.DIESEL | 0.0012 | -0.0958 | 0.0958 | 0.0000 | 0.0087 |
| CORN | 0.0006 | -0.2543 | 0.2028 | 0.0020 | 0.0422 |
| WHEAT | 0.0003 | -0.1699 | 0.1595 | -0.0009 | 0.0444 |
| SUGAR.CANE | 0.0010 | -0.1921 | 0.2008 | -0.0022 | 0.0460 |
| SUGAR.BEETS | 0.0008 | -0.1546 | 0.1151 | 0.0010 | 0.0364 |
| BR.SUGAR | 0.0012 | -0.1954 | 0.1411 | 0.0000 | 0.0343 |
| PALM.OIL | 0.0002 | -0.1539 | 0.1363 | 0.0005 | 0.0331 |
| RAPESEED | 0.0003 | -0.1289 | 0.0792 | 0.0021 | 0.0252 |
| SOYBEANS | 0.0002 | -0.2989 | 0.2333 | 0.0032 | 0.0400 |
| SUNFLOWERSEED | 0.0006 | -0.1972 | 0.1450 | 0.0025 | 0.0341 |
| COCOA | 0.0005 | -0.2336 | 0.1155 | 0.0014 | 0.0412 |
| COFFEE | 0.0007 | -0.1450 | 0.1863 | 0.0000 | 0.0416 |
| RICE | 0.0003 | -0.2312 | 0.1252 | 0.0008 | 0.0364 |
| ORANGES | 0.0007 | -0.1947 | 0.2072 | 0.0005 | 0.0489 |
| CATTLE | 0.0005 | -0.1122 | 0.0944 | 0.0009 | 0.0239 |
| COTTON | -0.0000 | -0.1468 | 0.1690 | -0.0004 | 0.0378 |

| | | | | | |
|---|---|---|---|---|---|
| BOVESPA | 0.0020 | -0.2233 | 0.1684 | 0.0045 | 0.0350 |
| DAX | 0.0014 | -0.2435 | 0.1494 | 0.0042 | 0.0290 |
| DOW.JONES | 0.0012 | -0.2003 | 0.1070 | 0.0029 | 0.0221 |
| FTSE.100 | 0.0006 | -0.2363 | 0.1258 | 0.0021 | 0.0234 |
| S.P.500 | 0.0012 | -0.2008 | 0.1136 | 0.0023 | 0.0231 |
| BRL.USD | 0.0003 | -0.0541 | 0.1451 | -0.0011 | 0.0172 |
| USD.EUR | -0.0001 | -0.0438 | 0.0658 | 0.0000 | 0.0108 |
| LIBOR | 0.0010 | -0.2646 | 0.1692 | 0.0014 | 0.0333 |
| FED.FUNDS | 0.0011 | -1.3269 | 0.9719 | 0.0000 | 0.1210 |

Source: Author's computation

# Appendix B

## Stationarity and Normality tests

Table 3: Stationarity tests

| | ADF value | p-value | KPSS value | p-value |
|---|---|---|---|---|
| BOVESPA | -9.162 | 0.01 | 0.1321 | 0.10 |
| BR.DIESEL | -9.272 | 0.01 | 0.1045 | 0.10 |
| BR.ETHANOL | -9.238 | 0.01 | 0.0927 | 0.10 |
| BR.GASOLINE | -8.910 | 0.01 | 0.0908 | 0.10 |
| BR.SUGAR | -9.424 | 0.01 | 0.2095 | 0.10 |
| BRENT.CRUDE | -7.751 | 0.01 | 0.1659 | 0.10 |
| BRL.USD | -8.063 | 0.01 | 0.3834 | 0.08 |
| CATTLE | -8.384 | 0.01 | 0.0837 | 0.10 |
| COCOA | -9.240 | 0.01 | 0.1160 | 0.10 |
| COFFEE | -8.873 | 0.01 | 0.3054 | 0.10 |
| CORN | -8.174 | 0.01 | 0.1110 | 0.10 |
| COTTON | -7.910 | 0.01 | 0.0483 | 0.10 |
| DAX | -9.920 | 0.01 | 0.0703 | 0.10 |
| DOW.JONES | -9.625 | 0.01 | 0.1424 | 0.10 |
| EU.BIODIESEL | -7.853 | 0.01 | 0.0889 | 0.10 |
| FED.FUNDS | -7.502 | 0.01 | 0.3750 | 0.09 |
| FTSE.100 | -10.342 | 0.01 | 0.0574 | 0.10 |
| GE.DIESEL | -8.498 | 0.01 | 0.1826 | 0.10 |
| GE.GASOLINE | -8.700 | 0.01 | 0.1052 | 0.10 |
| HEATING.OIL | -7.753 | 0.01 | 0.1658 | 0.10 |
| LIBOR | -6.867 | 0.01 | 0.3760 | 0.09 |
| NATURAL.GAS | -9.066 | 0.01 | 0.0545 | 0.10 |
| ORANGES | -8.878 | 0.01 | 0.1136 | 0.10 |
| PALM.OIL | -8.206 | 0.01 | 0.0624 | 0.10 |
| RAPESEED | -8.352 | 0.01 | 0.0677 | 0.10 |

| | | | | |
|---|---|---|---|---|
| RICE | -8.517 | 0.01 | 0.1093 | 0.10 |
| S.P.500 | -9.391 | 0.01 | 0.1139 | 0.10 |
| SOYBEANS | -7.852 | 0.01 | 0.0786 | 0.10 |
| SUGAR.BEETS | -9.108 | 0.01 | 0.2813 | 0.10 |
| SUGAR.CANE | -8.985 | 0.01 | 0.1769 | 0.10 |
| SUNFLOWERSEED | -7.905 | 0.01 | 0.0367 | 0.10 |
| US.BIODIESEL | -7.658 | 0.01 | 0.1300 | 0.10 |
| US.DIESEL | -7.254 | 0.01 | 0.1944 | 0.10 |
| US.ETHANOL | -10.224 | 0.01 | 0.0618 | 0.10 |
| US.GASOLINE | -8.067 | 0.01 | 0.1160 | 0.10 |
| USD.EUR | -9.042 | 0.01 | 0.1080 | 0.10 |
| WHEAT | -9.865 | 0.01 | 0.1013 | 0.10 |
| WTI | -7.410 | 0.01 | 0.1392 | 0.10 |

Source: Author's computation

Table 4: Normality tests

| | **Shapiro-Wilk** | **p-value** | **Jarque-Bera** | **p-value** |
|---|---|---|---|---|
| BOVESPA | 0.963 | 0.00 | 529.7818 | $< 0.01$ |
| BR.DIESEL | 0.386 | 0.00 | 97526.3395 | 0.00 |
| BR.ETHANOL | 0.885 | 0.00 | 3115.2708 | 0.00 |
| BR.GASOLINE | 0.540 | 0.00 | 62392.3730 | 0.00 |
| BR.SUGAR | 0.963 | 0.00 | 330.2356 | 0.00 |
| BRENT.CRUDE | 0.964 | 0.00 | 480.9791 | 0.00 |
| BRL.USD | 0.937 | 0.00 | 1872.1255 | 0.00 |
| CATTLE | 0.980 | 0.00 | 118.5170 | 0.00 |
| COCOA | 0.990 | 0.00 | 57.7288 | 0.00 |
| COFFEE | 0.993 | 0.00 | 27.4259 | 0.00 |
| CORN | 0.963 | 0.00 | 400.0296 | 0.00 |
| COTTON | 0.979 | 0.00 | 121.0067 | 0.00 |
| DAX | 0.928 | 0.00 | 2542.4433 | 0.00 |
| DOW.JONES | 0.922 | 0.00 | 3509.0887 | 0.00 |
| EU.BIODIESEL | 0.952 | 0.00 | 674.6624 | 0.00 |
| FED.FUNDS | 0.657 | 0.00 | 52118.7436 | 0.00 |
| FTSE.100 | 0.890 | 0.00 | 8711.2699 | 0.00 |
| GE.DIESEL | 0.964 | 0.00 | 285.8891 | 0.00 |
| GE.GASOLINE | 0.926 | 0.00 | 1004.7170 | $< 0.01$ |
| HEATING.OIL | 0.984 | 0.00 | 98.3155 | $< 0.01$ |
| LIBOR | 0.728 | 0.00 | 9537.0722 | $< 0.01$ |
| NATURAL.GAS | 0.991 | 0.00 | 34.3560 | 0.00 |
| ORANGES | 0.985 | 0.00 | 72.5479 | 0.00 |
| PALM.OIL | 0.984 | 0.00 | 97.6976 | 0.00 |
| RAPESEED | 0.946 | 0.00 | 524.2046 | 0.00 |
| RICE | 0.977 | 0.00 | 255.8161 | 0.00 |
| S.P.500 | 0.912 | 0.00 | 3174.2968 | 0.00 |
| SOYBEANS | 0.912 | 0.00 | 2477.9392 | 0.00 |
| SUGAR.BEETS | 0.990 | 0.00 | 37.1346 | 0.00 |

| | | | | |
|---|---|---|---|---|
| SUGAR.CANE | 0.986 | 0.00 | 68.1050 | 0.00 |
| SUNFLOWERSEED | 0.968 | 0.00 | 262.1125 | 0.00 |
| US.BIODIESEL | 0.958 | 0.00 | 472.6451 | 0.00 |
| US.DIESEL | 0.916 | 0.00 | 1634.7338 | 0.00 |
| US.ETHANOL | 0.907 | 0.00 | 2479.6674 | 0.00 |
| US.GASOLINE | 0.925 | 0.00 | 2248.6677 | 0.00 |
| USD.EUR | 0.981 | 0.00 | 175.3539 | 0.00 |
| WHEAT | 0.991 | 0.00 | 29.4118 | 0.00 |
| WTI | 0.954 | 0.00 | 753.7008 | 0.00 |

Source: Author's computation

# Appendix C

## The links in the individual periods

Table 5: The links of the MST in the entire period

| Variable 1 | Variable 2 | Distance metric |
|---|---|---|
| DOW.JONES | S.P.500 | 0.228 |
| US.BIODIESEL | EU.BIODIESEL | 0.386 |
| WTI | BRENT.CRUDE | 0.506 |
| HEATING.OIL | BRENT.CRUDE | 0.517 |
| DAX | FTSE.100 | 0.535 |
| FTSE.100 | S.P.500 | 0.579 |
| GE.GASOLINE | GE.DIESEL | 0.625 |
| US.DIESEL | US.GASOLINE | 0.664 |
| SUGAR.BEETS | SUGAR.CANE | 0.694 |
| BR.GASOLINE | BR.DIESEL | 0.791 |
| S.P.500 | BOVESPA | 0.825 |
| WHEAT | CORN | 0.863 |
| EU.BIODIESEL | RAPESEED | 0.960 |
| CORN | SOYBEANS | 0.971 |
| US.GASOLINE | GE.GASOLINE | 0.998 |
| RAPESEED | SOYBEANS | 1.057 |
| EU.BIODIESEL | PALM.OIL | 1.111 |
| BR.SUGAR | BOVESPA | 1.111 |
| BOVESPA | BRENT.CRUDE | 1.115 |
| BR.ETHANOL | BR.SUGAR | 1.119 |
| HEATING.OIL | NATURAL.GAS | 1.172 |
| SUGAR.CANE | COFFEE | 1.183 |
| RICE | SOYBEANS | 1.194 |
| BOVESPA | COFFEE | 1.195 |
| RAPESEED | BRENT.CRUDE | 1.203 |
| WHEAT | COTTON | 1.208 |

| | | |
|---|---|---|
| CORN | US.ETHANOL | 1.210 |
| GE.DIESEL | USD.EUR | 1.216 |
| BRENT.CRUDE | COCOA | 1.217 |
| BOVESPA | USD.EUR | 1.238 |
| LIBOR | FED.FUNDS | 1.260 |
| US.DIESEL | BR.DIESEL | 1.289 |
| CATTLE | FTSE.100 | 1.291 |
| RAPESEED | ORANGES | 1.299 |
| RAPESEED | SUNFLOWERSEED | 1.303 |
| US.GASOLINE | FED.FUNDS | 1.332 |
| BRL.USD | LIBOR | 1.334 |

Source: Author's computation

Table 6: The links of the MST in the Period I

| **Variable 1** | **Variable 2** | **Distance metric** |
|---|---|---|
| DOW.JONES | S.P.500 | 0.278 |
| WTI | BRENT.CRUDE | 0.420 |
| BR.GASOLINE | BR.DIESEL | 0.523 |
| BRENT.CRUDE | HEATING.OIL | 0.573 |
| GE.GASOLINE | GE.DIESEL | 0.585 |
| FTSE.100 | DAX | 0.617 |
| US.DIESEL | US.GASOLINE | 0.686 |
| DAX | S.P.500 | 0.698 |
| S.P.500 | BOVESPA | 0.808 |
| SUGAR.BEETS | SUGAR.CANE | 0.825 |
| US.GASOLINE | GE.GASOLINE | 0.918 |
| WHEAT | CORN | 0.945 |
| NATURAL.GAS | HEATING.OIL | 1.051 |
| CORN | SOYBEANS | 1.066 |
| SOYBEANS | RAPESEED | 1.130 |
| RAPESEED | EU.BIODIESEL | 1.136 |
| RAPESEED | PALM.OIL | 1.142 |
| BR.SUGAR | BOVESPA | 1.142 |
| GE.DIESEL | USD.EUR | 1.167 |
| BR.ETHANOL | BR.SUGAR | 1.179 |
| BOVESPA | COCOA | 1.204 |
| BOVESPA | COFFEE | 1.212 |
| EU.BIODIESEL | RICE | 1.221 |
| COTTON | RAPESEED | 1.222 |
| USD.EUR | COCOA | 1.230 |
| BRENT.CRUDE | SUGAR.CANE | 1.241 |
| BR.DIESEL | US.DIESEL | 1.250 |
| COFFEE | CORN | 1.255 |
| HEATING.OIL | COFFEE | 1.256 |

| | | |
|---|---|---|
| WTI | CATTLE | 1.263 |
| FED.FUNDS | ORANGES | 1.280 |
| US.ETHANOL | SUGAR.BEETS | 1.292 |
| SUGAR.CANE | ORANGES | 1.304 |
| PALM.OIL | SUNFLOWERSEED | 1.308 |
| FED.FUNDS | LIBOR | 1.308 |
| BRL.USD | LIBOR | 1.324 |

Source: Author's computation

Table 7: The links of the MST in the Period II

| Variable 1 | Variable 2 | Distance metric |
|---|---|---|
| DOW.JONES | S.P.500 | 0.185 |
| US.BIODIESEL | EU.BIODIESEL | 0.349 |
| DAX | FTSE.100 | 0.401 |
| BRENT.CRUDE | HEATING.OIL | 0.438 |
| FTSE.100 | S.P.500 | 0.464 |
| FTSE.100 | BOVESPA | 0.579 |
| WTI | BRENT.CRUDE | 0.593 |
| US.GASOLINE | US.DIESEL | 0.623 |
| SUGAR.CANE | SUGAR.BEETS | 0.669 |
| GE.GASOLINE | GE.DIESEL | 0.673 |
| EU.BIODIESEL | RAPESEED | 0.715 |
| WHEAT | CORN | 0.742 |
| EU.BIODIESEL | PALM.OIL | 0.853 |
| CORN | SOYBEANS | 0.884 |
| BRENT.CRUDE | BOVESPA | 0.928 |
| US.ETHANOL | CORN | 0.949 |
| BOVESPA | BR.SUGAR | 0.962 |
| RAPESEED | SOYBEANS | 0.974 |
| RAPESEED | HEATING.OIL | 0.990 |
| COCOA | COFFEE | 1.000 |
| US.DIESEL | GE.DIESEL | 1.003 |
| BRENT.CRUDE | COFFEE | 1.036 |
| COFFEE | COTTON | 1.064 |
| CORN | RICE | 1.082 |
| BRENT.CRUDE | SUGAR.BEETS | 1.083 |
| BR.SUGAR | BR.ETHANOL | 1.097 |
| DAX | CATTLE | 1.098 |
| BRENT.CRUDE | USD.EUR | 1.126 |
| HEATING.OIL | NATURAL.GAS | 1.182 |

| | | |
|---|---|---|
| RAPESEED | SUNFLOWERSEED | 1.208 |
| USD.EUR | GE.DIESEL | 1.211 |
| FED.FUNDS | US.GASOLINE | 1.215 |
| RAPESEED | ORANGES | 1.217 |
| FED.FUNDS | LIBOR | 1.258 |
| LIBOR | BRL.USD | 1.309 |
| BR.DIESEL | US.DIESEL | 1.313 |
| US.GASOLINE | BR.GASOLINE | 1.333 |

Source: Author's computation

Table 8: The links of the MST in the Period III

| Variable 1 | Variable 2 | Distance metric |
|---|---|---|
| DOW.JONES | S.P.500 | 0.266 |
| WTI | BRENT.CRUDE | 0.494 |
| BR.DIESEL | BR.GASOLINE | 0.550 |
| S.P.500 | FTSE.100 | 0.625 |
| US.BIODIESEL | EU.BIODIESEL | 0.656 |
| DAX | FTSE.100 | 0.659 |
| HEATING.OIL | BRENT.CRUDE | 0.661 |
| GE.GASOLINE | GE.DIESEL | 0.683 |
| SUGAR.BEETS | SUGAR.CANE | 0.701 |
| US.DIESEL | US.GASOLINE | 0.737 |
| EU.BIODIESEL | RAPESEED | 0.963 |
| US.GASOLINE | GE.GASOLINE | 0.984 |
| FTSE.100 | BOVESPA | 0.986 |
| WHEAT | CORN | 0.988 |
| CORN | SOYBEANS | 1.031 |
| RAPESEED | SOYBEANS | 1.038 |
| BR.ETHANOL | BR.SUGAR | 1.065 |
| BR.SUGAR | SUGAR.CANE | 1.097 |
| SUGAR.CANE | COFFEE | 1.109 |
| EU.BIODIESEL | PALM.OIL | 1.124 |
| BRENT.CRUDE | FTSE.100 | 1.127 |
| COTTON | US.BIODIESEL | 1.186 |
| BR.SUGAR | BOVESPA | 1.193 |
| CORN | US.ETHANOL | 1.198 |
| USD.EUR | US.BIODIESEL | 1.201 |
| GE.DIESEL | USD.EUR | 1.215 |
| WHEAT | COFFEE | 1.238 |
| WTI | COCOA | 1.248 |
| FED.FUNDS | LIBOR | 1.251 |

| | | |
|---|---|---|
| BR.ETHANOL | BR.GASOLINE | 1.252 |
| BRL.USD | SUNFLOWERSEED | 1.261 |
| USD.EUR | ORANGES | 1.268 |
| WTI | NATURAL.GAS | 1.279 |
| US.DIESEL | CATTLE | 1.288 |
| PALM.OIL | SUNFLOWERSEED | 1.295 |
| RICE | US.BIODIESEL | 1.297 |
| LIBOR | BRL.USD | 1.299 |

Source: Author's computation

Table 9: The links of the MST in the Period IV

| Variable 1 | Variable 2 | Distance metric |
|---|---|---|
| DOW.JONES | S.P.500 | 0.267 |
| WTI | BRENT.CRUDE | 0.356 |
| HEATING.OIL | BRENT.CRUDE | 0.372 |
| US.BIODIESEL | EU.BIODIESEL | 0.475 |
| SUGAR.BEETS | SUGAR.CANE | 0.527 |
| GE.GASOLINE | GE.DIESEL | 0.535 |
| US.DIESEL | US.GASOLINE | 0.564 |
| FTSE.100 | DAX | 0.755 |
| DAX | S.P.500 | 0.767 |
| WHEAT | CORN | 0.841 |
| BR.DIESEL | BR.GASOLINE | 0.881 |
| SOYBEANS | CORN | 0.944 |
| EU.BIODIESEL | RAPESEED | 1.061 |
| BR.SUGAR | BR.ETHANOL | 1.089 |
| S.P.500 | BOVESPA | 1.110 |
| GE.DIESEL | US.DIESEL | 1.114 |
| SUGAR.CANE | COFFEE | 1.142 |
| SUGAR.CANE | BR.ETHANOL | 1.149 |
| WTI | S.P.500 | 1.163 |
| CORN | US.ETHANOL | 1.173 |
| WTI | US.ETHANOL | 1.187 |
| BOVESPA | COFFEE | 1.190 |
| US.GASOLINE | BR.GASOLINE | 1.191 |
| US.BIODIESEL | BRENT.CRUDE | 1.201 |
| BR.GASOLINE | BR.ETHANOL | 1.206 |
| US.GASOLINE | ORANGES | 1.207 |
| US.BIODIESEL | PALM.OIL | 1.215 |
| CORN | RICE | 1.239 |
| HEATING.OIL | NATURAL.GAS | 1.242 |

| | | |
|---|---|---|
| WTI | COCOA | 1.253 |
| FTSE.100 | CATTLE | 1.258 |
| DAX | COTTON | 1.280 |
| GE.DIESEL | USD.EUR | 1.284 |
| US.ETHANOL | FED.FUNDS | 1.334 |
| COTTON | LIBOR | 1.342 |
| SUNFLOWERSEED | US.BIODIESEL | 1.351 |
| EU.BIODIESEL | BRL.USD | 1.356 |

Source: Author's computation