Universidad de San Andrés

Escuela de Administración y negocios

Maestría en Finanzas

Energy and vegetable oils : a price transmission network

Autor: Ahlsén Berg, Fredrik

Legajo: 56502200

Mentor: Cortina, Elsa

2015

Energy and Vegetable Oils

- A price transmission network

QUAERERE VERUM
Universidad de
SanAndres

Author:
Fredrik Ahlsén Berg
Universidad de San Andrés
Finanzas
fahlsenberg@gmail.com

Tutor:
Elsa Cortina
Universidad de San Andrés
Finanzas
Elsa.Cortina@gmail.com

Abstract

This thesis studies how energy and agricultural commodities are linked together and in particular how they form a network between each other. Cash traded commodities are substituted by their closest exchange traded alternative. The network is presented in the form of Minimal Spanning Trees and Hierarchical Trees. It is considered under various frequencies and periods that allow us to determine how the network is affected by those factors and what the implications are for a biofuels trader. It is shown that the networks form two clusters that are connected via soybean oil. An analysis is made on the effects of the food and financial crises on the energy and agricultural commodities used for the production of biodiesel. In this context, it is shown that the co-movement of energy and agricultural commodities increased during the crises period and that the network is denser after than before the crises. Finally, it is argued that substituting the cash markets with their closest exchange traded alternative does change the network's dynamics.

QUAERERE VERUM
Universidad de
SanAndrés

QUAERERE VERUM
Universidad de San Andrés

1 Introduction

This paper is motivated by the increased usage of biofuels for the world's energy supply and how this affects the price transmission network of regular food and energy commodities. The complex network of food and energy commodities is presented using methods of graph theory and taxonomy. Thus clearly identifying the relationships between components of the system without imposing the structural and distributional assumptions that other methods demand. The price transmission network is analysed under different market phases and time horizons. With these results, the effects of the introduction of biofuels to the price transmission between agricultural and energy commodities can be better understood.

The rapid, policy driven, expansion of the use of biofuels since the end of the 20th century has, in combination with high commodity prices during the period of 2007-2008, fuelled a debate about the effects of the introduction of biofuels that is polemic. Biofuels are produced from agricultural products and the economics of the industry is driven by both the price that the end product may fetch as well as the price of the feedstock to produce it. Consequently, biofuels can be seen to bridge the food and energy sectors. The scare of a socioeconomic effect of the increased importance of biofuels has spurred a vast amount of research to be done on the economics of biofuels.

The two articles by (Zilberman et al. 2012) and (Janda, Kristoufek and Zilberman 2012) summarize the latest findings from the research of the economics of biofuels. From this meta analysis it is clear that the simulation models, and partial and general equilibrium models rely on the assumption of a relationship between the prices of food, biofuels and energies. So far, the empirical research has been inconclusive as to this matter (Kristoufek, Janda and Zilberman 2012).

The currently published empirical papers have used many different methods to establish a statistical relationship between food, biofuels and energy. To name but a few, the spectrum spans from Markowitz portfolio theory (Zhang, Lohr and Wetzstein 2008), to various applications of co-integration (Serra, Zilberman and Gil 2011) and (Serra, Zilberman, et.al. 2011), to investigations of the volatility using co-integration, vector error correction model (VECM), and multivariate generalized autoregressive

Universidad de SanAndrés QUAERERE VERUM

conditional heteroskedasticity (MGARCH) models (Zhang, Lohr and Escalante, et.al. 2009), to volatility spillover using Bayesian Markov Chain Monte Carlo methods to estimate the model's parameters (Du, Yu and Hayes 2011). The methods that have been used are sophisticed and they impose structural and distributional assumptions between the prices of biofuels and related commodities that does impair their flexibility (Kristoufek, Janda and Zilberman 2012). In this paper, the problem of assumptions is completely circumvented by the use of Minimal Spanning Trees (MSTs) and Hierarchical Trees (HTs) to map the network of price transmissions.

MSTs and HTs have been widely used to map networks in distinct fields of science, everything from biology to physics. More recently, econophysicists have begun to apply the method in their field in order to map and understand the complex networks that they investigate. MSTs and HTs have for example been used to map the networks of: stocks connections (Bonanno, et.al. 2004), foreign exchange (Wang, et.al. 2012), term structure of interest rates (Tabak, Serra and Cajueiro 2009) and commodities (Tabak, Serra and Cajueiro 2010) .

In this paper, we apply MST and HT analysis to the network of biofuels and commodities that one would suspect are related to their production *i.e.* agricultural and energy commodities. The benefit of using MST and HT analysis is that we are able to simultaneously analyse a great number of commodities without blurring the picture by being too complex. In other words, even though the web may consist of many nodes one may identify clusters and individual connections clearly. Furthermore, bootstrapping the data indicates how strong the respective links are as well as the strength of the network on its own. This fundamental analysis may then serve as a spring board for further in-depth analysis of individual price connections using traditional econometrical methods such as co-integration and mean revertion.

Applying MSTs and HTs to the biofuels sector has been done before in (Kristoufek, Janda and Zilberman 2012). In their analysis they show that the two clusters are formed: one for bioethanol and another for biodiesel. This makes sense as bioethanol is used in blends with gasoline and biodiesel for blends with diesel and as such their demand is distinct. Not even the feedstock is the same; bioethanol is produced from agricultural commodities high in sugar content such as corn, sugar cane and wheat whereas biodiesel is mainly produced from vegetable oils. Here we will focus on the biodiesel space.

QUAERERE VERUM
Universidad de San Andrés

Biodiesel is produced from vegetable oils that can be derived from many types of feedstock. Some of the most common ones are rapeseed, soybean and oil palm from which Rape Methyl Ester (RME), Soy Methyl Ester (SME) and Palm Methyl Ester (PME) are produced. There are other variations as well, such as Fatty Acid Methyl Ester (FAME) and advanced biofuels but the original feedstock is vegetable oil.

These products are traded Over-the-Counter (OTC) and therefore the current literature on price transmission of biofuels, agricultural products and energy commodities, such as (Peri and Baldi 2013), (Peri and Baldi 2008), (Kristoufek, Janda, and Zilberman 2012) to name a few, all consider illiquid or untradeable price indices in the course of their analysis. By definition, such indices are not traded on an exchange.

What makes this paper different from previous research is that we substitute these illiquid indices/products with the closest related exchange traded commodity. As an example, MATIF Rapeseed and Winnipeg Canola substitute rapeseed and canola oil respectively as the oil is only traded in the cash market whereas the seed is traded on an exchange. By doing so we aim to understand the dynamics of the web of liquid tradable financial assets related to biodiesel.

In order to explore the network, we want to analyse the behaviour of this group of exchange-traded commodities under different time frequencies (daily, weekly and monthly) as well as before, during and after the financial and food crises. Through such thorough analysis we want to establish the network's short-/long-term structure as well as whether there have been any structural changes due to amended regulations and/or new technological advances. The hypothesis is that the substitution of untradeable indices and illiquid commodities by their closest liquid alternative does not change the dynamics of the network that represents the biodiesel space. This hypothesis is not confirmed as the network does not demonstrate the attributes that were expected. Instead we find that the food and energy space moved closer together during the crises period and that it has remained closer after the crises. Furthermore, we find that short-term shocks can cause temporary distortions to the network but that it reverts again in the medium- to long-term unless there has been a structural or macro-economical change that may be regulatory driven.

The rest of this paper is organized as follows. Section 2 includes a brief historical background to the biofuels industry as well as the presentation of a handful of papers that are important for this paper. In section 3 we account for the methods used in this paper. Section 4 introduces the data and includes a graphical presentation of the price series. The results are presented and analysed in section 5. Finally, section 6 concludes and discusses shortcomings as well as suggestions for future research.

2 Background and Previous Research

The usage of agricultural products as fuel is nothing new. On the contrary, the French Otto Company reportedly ran a diesel engine on peanut oil at the 1900 World's Fair in Paris. Allegedly, the French government wanted to find an energy source for their tropical colonies that was produced locally in order to avoid shipping in coal and liquid fuels. Research continued and in 1937 the Belgian G. Chavanne filed the patent 422.87 for what was probably the world's first biodiesel and in 1938 it was put to commercial use on a bus line between Brussels and Louvain (Leuven) with a satisfactory result (Knothe 2011). During World War II many countries experimented with the use of vegetable oils, as fuel was scarce due to the limited, and controlled, supply of mineral oil (Yergin 2008). After the war, mineral oil was made available in abundance and the interest in vegetable oil as fuel declined until the oil crisis of the 1970s when the technology was rediscovered (Knothe 2011).

The reintroduction of biofuels was led by ethanol inclusion in the US and Brazil during the 1970s. Ethanol subsidies were introduced in the US Energy Tax Act of 1978 as the US had 'the desire to reduce dependence on imported fossil fuels; to reduce greenhouse gas (GHG) emissions; and to increase demand for domestic farm commodities serving as raw material for biofuels.' Similar subsidies for biodiesel were only introduced with the Conservation Reauthorization Act in 1998 (Janda, Kristoufek, and Zilberman 2012).

The Energy Policy Act of 2005 under the Renewable Fuel Standard (RFS) introduced mandates expressed in volumetric terms. For 2006, the objective was to incorporate 4 billion gallons of renewable fuels, to be increased to 7.5 billion gallons by 2012. With the Energy Independence and Security Act of 2007 the RFS was expanded to reach 36 billion gallons by 2022 of which 1 billion gallons biodiesel (Trujillo-Barrera, Mallory, and Garcia 2012). It is this volumetric Ethanol and Biodiesel Excise Tax Credits that lately has provided the largest subsidies (Janda, Kristoufek, and Zilberman 2012).

The European bioenergy revolution started with the Common Agricultural Policy (CAP) in 1992. The CAP was an attempt to stop the overproduction of agricultural goods in Europe. As a part of this process, arable land was supposed to be converted to non-food use such as energy-crop production (Pelkmans et al. 2007).

Additionally, in the beginning of the 21st century the EU had a new energy agenda that promoted reduced greenhouse gas (GHG) emissions, energy security and the reduction of urban pollution while at the same time improving the financial situation for farms (OECD 2008). In order to achieve these goals EU decided to use both a blending mandate as well as a tax relief (Peri and Baldi 2013).

These incentives were first introduced with the 2003 Biofuels Directive (2003/30/EC) which mandated that 2% of all the transport fuel should come from renewable resources by 2005 and that it should gradually increase to 5.75% biofuels inclusion by 2010. Via the Energy Taxation Directive (2003/96/EC) Member States were allowed to promote biofuels inclusion through tax exemption under the conditions that:

- The tax exemption or reduction must not exceed the amount of taxation payable on the volume of renewable used;
- Change in the feedstock prices are accounted for in order to avoid overcompensation;
- The exemption or reduction authorized may not be applied for a period of more than six consecutive years. This is renewable.

(Pelkmans et al. 2007)

In the beginning the Fuel Quality Directive restricted the biofuels inclusion to 5% for both bioethanol and biodiesel (Pelkmans et al. 2007). In 2009, the Renewable Energy Directive (R.E.D.) (2009/28/EC) raised the bar by setting 10% biofuels inclusion as a goal for 2020.

At the same time, the EU continued to work on how to improve the farmers' finances by developing the CAP. In 2004, a EUR 45/ha incentive was introduced to induce farmers to grow energy-crops all while there also existed another payment for cultivating non-food crops on set-aside land.

With the financial crisis the objectives of the EU changed and in 2009 both incentives were abolished following the 2007 Health Check reform (Peri and Baldi 2013).

Before their abolition, the above mentioned political subsidies led to artificial high profitability in the biodiesel sector during the period from the Biofuels Directive in 2003 till 2006, when many tax incentives were abandoned in favour of blending mandates. As a textbook example of supply and stimulated demand, the high profits attracted new entrants. Which, in turn, led to an explosive expansion of the production capacity in Europe and today the EU is the world's largest biodiesel producer (Jung et al. 2010).

Figure 1, on the next page, illustrates the relationship between the expansion of production and the increase in demand for the period 2003-2011 (European Biodiesel Board 2014). One can see that the addition of production capacity outpaced the increase in demand ever since 2006 and that the decrease in utilization was drastic between 2006 and 2007.

It has been argued that the rapid production decrease from 2006 to 2007 was a consequence of a change from tax incentives to mandatory blending requirements. These changes, effectively, erased much of the regulatory protection that producers had enjoyed up until then (Peri and Baldi 2013). Since this policy change, the European biodiesel industry has been burdened with overcapacity that has kept returns oscillating from negative to slightly positive even though the yearly production increased every year until 2010.

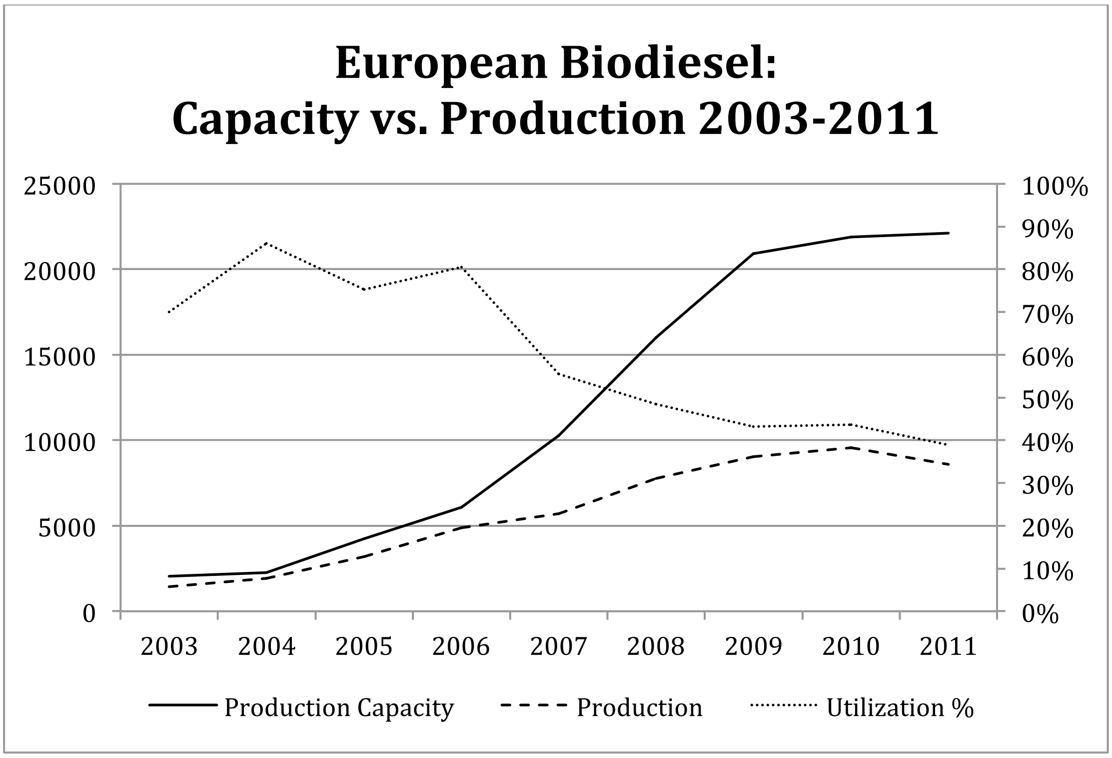

Figure 1 : European Biodiesel : Production vs. Capacity 2003-2011 (European Biodiesel Board 2014)

Worldwide, this continuous increase in regulatory demand for biofuels created additional demand for agricultural commodities. We will concentrate on biodiesel, which is produced from vegetable oils. Vegetable oils are a source of the essential nutrient fat. As such, vegetable oils are an important food staple that should be readily available to the world's population. The additional demand from the energy sector has arguably changed the relationship between vegetable oils and crude oil and its derivatives (Peri and Baldi 2013).

The additional demand from the biofuels sector has spawned a vibrant food vs. fuel debate, as there is concern that additional demand from the energy sector has inflated the food prices and made food increasingly difficult to obtain for the poor. Therefore, there is an abundance of literature available which either focus on the relationship between the food and fuel prices or concentrates on the impact of the introduction of biofuel on commodity food prices (Zilberman et al. 2012). The latter perspective is more political in the sense that it considers what the effects are of the introduction of biofuels on a socioeconomic level. This is a field on its own and of great interest to politicians but not for traders and will not be considered here. A handful of relevant research is presented below. For a thorough meta analysis, please refer to (Zilberman et al 2012) and (Janda, Kristoufek and Zilberman 2012).

(Kristoufek, Janda, and Zilberman 2013) set the theoretical framework for this article. They show that the biodiesel market can be considered as a market where the price is determined by the supply and demand of the commodity with technological and regulatory constraints that distort the prices and consequently the economics of the industry. From this theoretical framework they analyse the price transmission between biofuels and related commodities. The authors find that both ethanol and biodiesel prices are responsive to their production factors and that the food and financial crises significantly increased the strength of this transmission.

(Peri and Baldi 2013) combines a multiple structural change approach with rolling co-integration. The authors analysed the long-run price relationship between rapeseed oil and fossil diesel prices on the FOB Amsterdam-Rotterdam-Antwerp (ARA) market, which is the most liquid cash market in Europe, for structural breaks over the period 1 January 2001 to 27 April 2010. They identified four structural breaks, all of which were closely related in time to either policy changes or macro-economical events. The authors conclude that rapeseed oil prices are affected by policy events and that the rapeseed oil prices rapidly converge to the economics driven by the fossil diesel prices. Finally, they note that the introduction of such a link may pose future problems for the farmers, as they may not be used to the volatility of the energy sector.

(Peri and Baldi 2008) used a threshold vector error correction model to investigate whether an asymmetric dynamic adjustment process existed among: rapeseed oil, sunflower oil, soybean oil and diesel prices. The authors used weekly data from January 2005 to November 2007 and all prices were Amsterdam-Rotterdam-Antwerp. They found that sunflower oil and soybean oil were not influenced by diesel prices during the period. Rapeseed oil on the other hand was found to have a strong link with diesel. They suggest that this was due to the high quota of EU biodiesel produced from rapeseed oil (80%) during the period.

The above works are mainly concerned with econometrics and price relationships. Another method that has been used is that of volatility spill over but so far only in Brazil and the United States when analysing biofuels (Zilberman et al. 2012).

As an example, (Trujillo-Barrera, Mallory, and Garcia 2012) analyses volatility spill over in the United States from crude oil to ethanol and corn. The authors use a trivariate model to find volatility linkages from crude oil to corn and ethanol during 2006-2011. This was a time when about 25-35% of all corn usage went to the ethanol industry. Their findings imply that the volatility of energy commodities spill over into the agricultural market once a certain threshold of the supply is used for energy markets. The effect thereof is found to be 15% on average and that it peaked at 45% when volatility was high. The authors point out that this may cause concerns for the risk management of agricultural commodities.

(Kristoufek, Janda, and Zilberman 2012) applied graph theory to a number of North American and European commodities that they deemed interesting for the analysis of the relationship between energy and agricultural commodities. The authors find that the analysis is almost inconclusive in the short-term but that it provides meaningful insights in the medium- to long-term as two separate branches form. They find that the food and financial crises altered the relationship between the commodities analysed.

Previous research found that the introduction of biofuels affect the commodity markets (Zilberman et al. 2012) and some papers, such as (Peri and Baldi 2013) and (Kristoufek, Janda, and Zilberman 2013), have found that there is a price transmission from energy to agricultural commodities. What makes this paper different is that it explores the possibility to substitute the cash markets with futures. This ensures that the network of commodities is tradable and by doing so we explore the commodities interchangeability.

3 Methodology

There is a three-step process to generate minimal spanning trees (MST) and hierarchical trees (HT) consisting of concerned commodities: First, we need to define a way to map the correlations to a suitable measure space. Secondly, we must find a way to construct the trees. Finally, we must verify that the links we found are sufficiently stable to be used in our work. With this in hand, we can set out to analyse the material.

The method we use was outlined in (Kristoufek, Janda, and Zilberman 2012). Part of the method was originally introduced in (Mantegna 1999) and the additional test of the strength of links was added in (Tumminello et al. 2007).

3.1 Mapping the Correlation Matrix to the Distance Matrix

In order to graphically express the interconnections between our time-series, we must find a suitable topological space. Normally, interconnections between various time-series are represented using a correlation matrix $\mathbb{C}$. $\mathbb{C}$ does not possess the properties that we need to express the interconnections graphically. To resolve this we construct a measure d_{ij} that will map $\mathbb{C}$ to another distance space and the distance matrix $\mathbb{D}$. In the distance space, the distance between the nodes represent the level of correlation. In this process, we first define the correlation matrix $\mathbb{C}$ and discuss some important properties. We then proceed to define what a measure is and construct the measure that will map $\mathbb{C}$ to $\mathbb{D}$. Finally, we discuss the properties of the measure d_{ij}.

3.1.1 Correlation Matrix

For a set S, with cardinality $|S| = n$, of weak stationary time-series *i.e.* all possess well-defined means and variances. The correlation matrix $\mathbb{C}$ consist of the sample correlations of all possible combinations of time-series. To formalize, for a pair of time-series i and j with respective values X_{it} and X_{jt} and $t = 1, \dots, T$. Define the sample correlation coefficient $\widehat{\rho_{ij}}$ as:

$$\widehat{\rho_{ij}} = \frac{\sum_{t=1}^{T}(X_{it} - \overline{X_i})(X_{jt} - \overline{X_j})}{\sqrt{\sum_{t=1}^{T}(X_{it} - \overline{X_i})^2 \sum_{t=1}^{T}(X_{jt} - \overline{X_j})^2}}, \tag{1}$$

where $\overline{X_i} = \frac{\sum_{t=1}^{T} X_{it}}{T}$ and $\overline{X_j} = \frac{\sum_{t=1}^{T} X_{jt}}{T}$ are the respective time-series averages (Kristoufek, Janda, and Zilberman 2012). As is usual, $\mathbb{C}$ consist of $n(n-1)/2$ pairs of correlation where $\widehat{\rho_{ij}} \in \{x : \mathbb{R} \wedge -1 \leq x \leq 1\}$. A value of $\widehat{\rho_{ij}} = -1$ represents perfect anti-correlation, $\widehat{\rho_{ij}} = 1$ perfect correlation, and $\widehat{\rho_{ij}} = 0$ no correlation whatsoever.

3.1.2 Distance Measure

Logically, the measure we define should map highly correlated pairs as short distances and anti-correlated pairs as long distances. Most importantly this distance space should be a sub-space of $\mathbb{R}^{+} \cup \{0\}$ as negative distances are contradictory. Define the mapping:

$$d_{ij} = \sqrt{2(1-\rho_{ij})} \qquad (2)$$

We know from before that $\widehat{\rho_{ij}} \in \{x: \mathbb{R} \wedge -1 \leq x \leq 1\}$. Consequently, $d_{ij} \in \{x: \mathbb{R} \wedge 0 \leq x \leq 2\}$ which is a sub-space of $\mathbb{R}^{+} \cup \{0\}$. In particular, $d_{ij} \to 0$ as $\widehat{\rho_{ij}} \to 1$ and $d_{ij} \to 2$ as $\widehat{\rho_{ij}} \to -1$. In other words: the stronger the correlation, the shorter the distance and d_{ij} fulfils the first part of our pre-requisites. It remains to be proven that d_{ij} is a measure.

The mapping d_{ij} is a measure if and only if:

$$\begin{array}{rl} \text{i.} & d_{ij} = 0 \text{ if and only if } i = j; \\ \text{ii.} & d_{ij} = d_{ji}; \\ \text{iii.} & d_{ij} \leq d_{ik} + d_{kj}, \forall k \in t. \end{array} \qquad (3)$$

<u>Proof that d is a measure</u> (Mantegna 1999)**:**

i. $(by\ contradiction)$: $d_{ij} \neq 0 \Rightarrow \widehat{\rho_{ij}} \in \{x: \mathbb{R} \wedge -1 \leq x < 1\} \Rightarrow i \neq j$

$i = j \Rightarrow \rho_{ij} = 1 \Rightarrow d_{ij} = 0$

ii. The correlation coefficient matrix $\mathbb{C}$ is symmetric by definition $\Rightarrow$ the distance matrix $\mathbb{D}$ is symmetric.

iii. Consider Euclidean distance between two vectors $\widetilde{Y_i}$ and $\widetilde{Y_j}$ whose components are the time-series values of the time-series Y_i and Y_j. The vector's unitary norm is constructed by subtracting to each record the average value of and by normalizing it to its standard deviation.

$$Q.E.D.$$

3.2 Minimal spanning tree and hierarchical tree

The construction of the minimal spanning tree (MST) will be done following the method outlined in (Mantegna 1999). The MST then serves as a mapping to map the distance matrix $\mathbb{D}$ to the subdominant ultrametric space and the matrix $\mathbb{D}^{<}$. In turn, the subdominant ultrametric matrix $\mathbb{D}^{<}$ forms the foundation for the hierarchical tree (HT).

3.2.1 Constructing the minimal spanning tree

The MST is derived from the Euclidean distance matrix $\mathbb{D}$. As a first step, discard the diagonal elements that are zero by definition. We use Kruskal's algorithm, outlined below, to sort and graph the remaining elements of $\mathbb{D}$.

> "Perform the following step as many times as possible: Among the [distances] of [$\mathbb{D}$] not yet chosen, choose the shortest [distance] which does not form any loops with those edges already chosen."
>
> (Kruskal 1956)

Kruskal proved that this is the minimal spanning tree for $\mathbb{D}$ and that it will have $n - 1$ connections instead of $n(n-1)/2$ that we had in the correlation matrix. This filtering results in a graph that is easier to read and analyze.

To illustrate the procedure, consider the distance matrix $\mathbb{D}_{example}$ that reflects the connections between the time-series in the set $\{A, B, C, D\}$. The respective distance metrics are: $d_{AB} = 0.1, d_{BC} = 0.2, d_{AC} = 0.5, d_{AD} = 0.6, d_{CD} = 0.8, d_{BD} = 1$. The shortest distance is $d_{AB} = 0.1$ and we adjoin A and B by this edge. The second shortest edge is $d_{BC} = 0.2$ and we adjoin C to the already existing growing tree by B. The third shortest edge is $d_{AC} = 0.5$ but if we were to add this edge we will have a loop so it is ignored. Then we follow this procedure until we have connected the n nodes as can be seen in Figure 2 below.

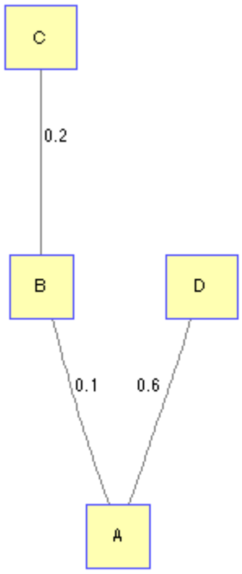

Figure 2 : Example MST

3.2.2 Constructing the hierarchical tree

The HT is constructed from the MST by mapping the distance matrix $\mathbb{D}$ to the subdominant ultrametric matrix $\mathbb{D}^{<}$ according to Mantegna's method.

> "Let $d_{ij}^{<}$ be the subdominant ultrametric distance between i and j as the maximum value of any Euclidean distance d_{kl} detected by moving in single steps from i to j through the shortest path connecting i and j in the MST."
>
> (Mantegna 1999)

To illustrate we use the example matrix $\mathbb{D}_{example}$ and map it using $d_{ij}^{<}$ to $\mathbb{D}^{<}$. Consider the connection from A to C, this will have the distance $d^{<} = 0.2$ as it is the largest edge that needs to be traversed in the associated MST. At the same time, the Euclidean distance is 0.5 but this distance is ignored in the construction of HT. In Figure 3 you see HT for the example matrix.

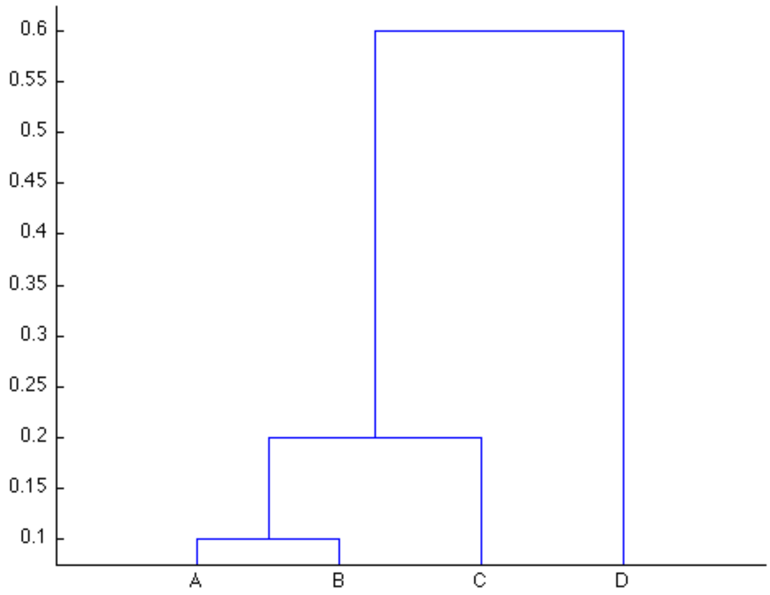

Figure 3 : HT Example

3.3 Stability of links

One of the problems when we construct the MST from a sample of time-series of finite length is that there is uncertainty as to whether the links are statistically significant or not (Tumminello et al. 2008). One way to circumvent this problem is to apply the bootstrap technique invented by (Efron 1979) and then popularized as a phylogenetic hierarchical tree evaluation method in (Felsenstein 1985). This method of so-called *bootstrap weights* of the respective vertices gives us guidance to both the global reliability of the graph as well as that of the individual edge (Tumminello et al. 2008).

3.3.1 Bootstrap - a way to measure link reliability

Consider a matrix $\boldsymbol{X}$ consisting of n time-series all of the length T that are assumed to be weak stationary. So $\boldsymbol{X}$ is a $n \times T$ matrix. From this matrix we derive the correlation matrix $\mathbb{C}$ that serves as the foundation for the construction of the MST. To ensure ourselves that the results we achieve are not mere statistical arbitrations we apply bootstrap procedure. Let us run a number of $r = 1000$ replicas of the original matrix $\boldsymbol{X}$. Each replica $\boldsymbol{X}_i^*$ is constructed by randomly selecting rows of $\boldsymbol{X}$ until we have achieved a total of T whilst allowing for repetition. In other words, some rows of $\boldsymbol{X}$ may be repeated whereas others are absent in the replica $\boldsymbol{X}_i^*$. For each replica $\boldsymbol{X}_i^*$, the associated correlation matrix $\mathbb{C}_i^*$ is found and the MST is calculated. The result is a set consisting of r MSTs $\{MST_1^*, \dots , MST_r^*\}$. This set of MSTs is used to find the bootstrap value for each link.

The bootstrap value is defined as the normalized recurrence of one specific link from the original MST in the group of replicated MSTs. To illustrate, say that nodes A and B are connected in the original MST. In the collection of replicated MSTs we find it in 750 out of 1000 cases. Then the bootstrap value is 0.750. It is this normalized value that will allow us to decide the link's stability.

By using the average of all the bootstrap values we can assess the reliability of the entire graph. Individual bootstrap values may allow us to identify communities within the graph. This technique does not require any knowledge of the data distribution. As such, it suits work with high dimensional systems where it can be difficult to infer the joint probability distribution (Tumminello et al. 2008).

The error associated with the correlation coefficient ρ roughly scales like $(1-\rho^2)/\sqrt{T}$ for normally distributed random variables. This implies that the error decreases as the correlation increases. However, this does not imply that the reliability of the link behaves accordingly (Tumminello et al. 2008).

4 Data

In this section we present the sets of commodities analyzed in this study. In particular, we present why these commodities were picked and how the data was gathered.

4.1 Commodity selection

The commodities analyzed in this study were picked as they represent viable links between Energies and Vegetable Oils. In order to provide liquidity only exchange-traded commodities were used, as the cash markets are non-accessible for most market participants.* To minimize the influence of adverse price movements that may be seen during the delivery period, the 2nd month futures were used. This will also ensure a liquid and tradable market.

* The cash market is the OTC basis/premium market that trade between commodity traders. The market is non-regulated and the counter-party risk remains with the parties.

The cross-section that we analyze consist of the daily, weekly and monthly prices[†] of Crude Oil-Brent (*Brent*), Crude Oil-WTI (*WTI*), Gasoil, Heating Oil (*HO*), Palm Oil (*CPO*), Soybean Oil (*SBO*), Canola seed (*Canola*) and Rapeseed for the period 13 Jan 2005-13 Jan 2014. The prices were pulled from Bloomberg and their respective exchanges, tickers and contract months are specified in Table 1.

Commodity	**Exchange**	**Ticker**	**Contract type**
Crude Oil - Brent	ICE	CO 2	2nd month futures
Crude Oil – WTI	NYMEX	CL 2	2nd month futures
Gasoil	ICE	QS 2	2nd month futures
Heating Oil	NYMEX	HO 2	2nd month futures
Palm Oil	BMD	PO 2	2nd month futures
Soybean Oil	CBOT	BO 2	2nd month futures
Canola seed	ICE	RS 2	2nd month futures
Rapeseed	LIFFE – Paris	IJ 2	2nd month futures

Table 1: Selection of commodities analyzed.

Canola seed and Rapeseed substitute Canola Oil and Rapeseed Oil (*RSO*) respectively. Substituting the oil with the seed in an analysis of the returns is logical thanks to the economics of Canola seed/Rapeseed processing. When crushing Canola seed/Rapeseed, approximately 40-42% of the seed's original weight is extracted as oil (The IntercontinentalExchange 2014). Typically, the value of the oil is much higher than that of the mid-protein that is the residual of the seed. Consequently, the bulk of the seed value is derived from the oil and the processors can be said to crush the seed for its oil. In other words, the crush margin depends to a great extent on the value of the oil and a strong link between the cash markets for oil and the exchange-traded seed should be expected in the deferred months.

[†] The 212 trading days where one or more of the commodities did not trade were ignored.

4.2 Prices and Spreads

A graphical analysis of the charts shown in Figure 4, Figure 5 and Figure 6 indicates that all of the commodities' prices moved in similar patterns during the period analysed. In particular, three sub-periods are identifiable with a distinctive break caused by the combination of the food and financial crises that struck the World 2007-2010. These sub-periods are here identified as: 1) the pre-food/financial crises (13 Jan 2005-9 Jul 2007), 2) the crises period (10 Jul 2007-26 Jan 2010) and 3) the post-crises period (27 Jan 2010-13 Jan 2013). The crises period starts with the food crisis that has been said to commence on the 9 Jul 2007 (Kristoufek, Janda, and Zilberman 2012) and it ends with President Barack Obama proclaiming the end of the financial crisis on the 26 Jan 2010 (United States Department of the Treasury 2010).

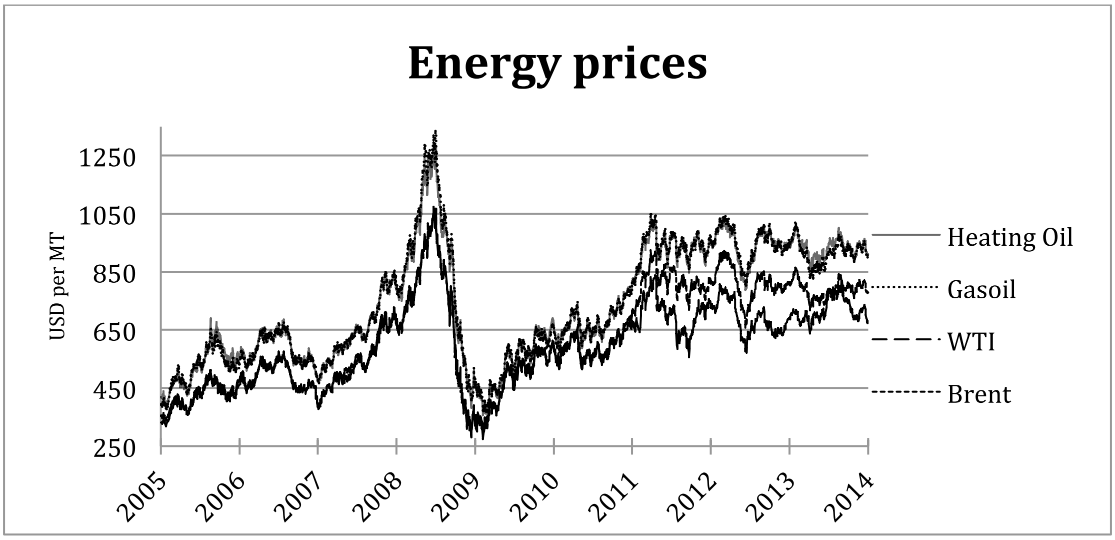

Figure 4 : USD per MT prices of Gasoil and Heating Oil

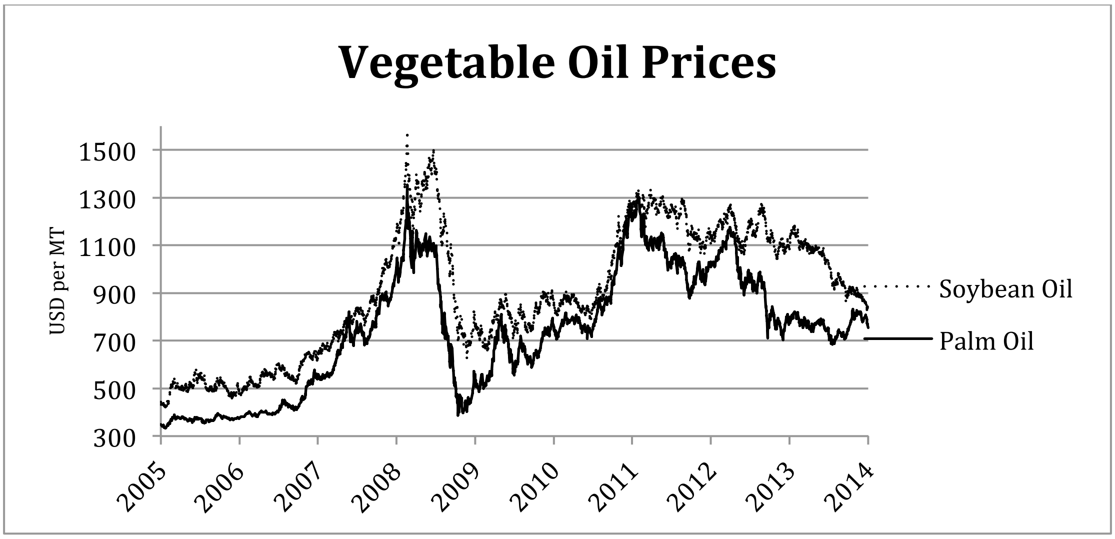

Figure 5 : SBO and CPO prices

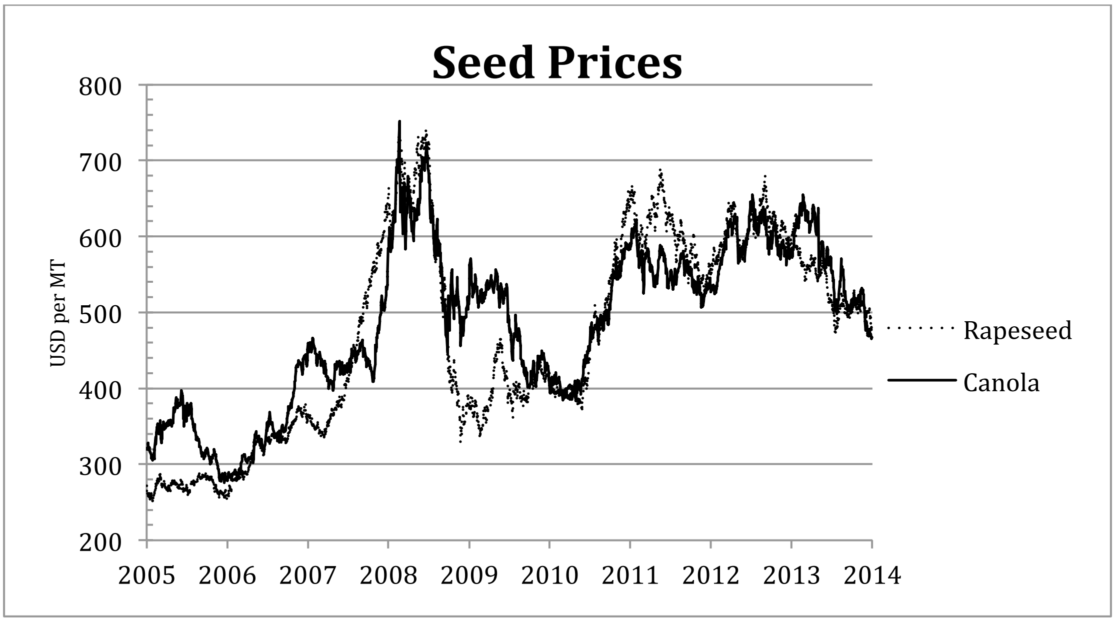

Figure 6 : Canola and Rapeseed Prices

During the pre-crises period prices increased with some minor corrections, once the food crisis or, perhaps better thought of as, commodity crisis commenced during 2007 the prices escalated fast and the correction during the financial crisis proved equally volatile. After the crises, the prices of the commodities have stabilized as well as the spreads between them.

By analysing the spreads between commodities one can better understand the supply and demand differences between associated commodities that are present in the market. For example, the spread between *Brent* and *WTI* represent the International and North American crude oil market respectively. The spread between the two commodities has changed dramatically during the period analysed thanks to the shale oil revolution in the United States (see Figure 7). All other spreads analysed have kept their dynamics intact. As an example, consider the *HO-Gasoil* spread in Figure 8. The construction of the MSTs and HTs is a way to identify which of the spreads in network of commodities may be worth to study further.

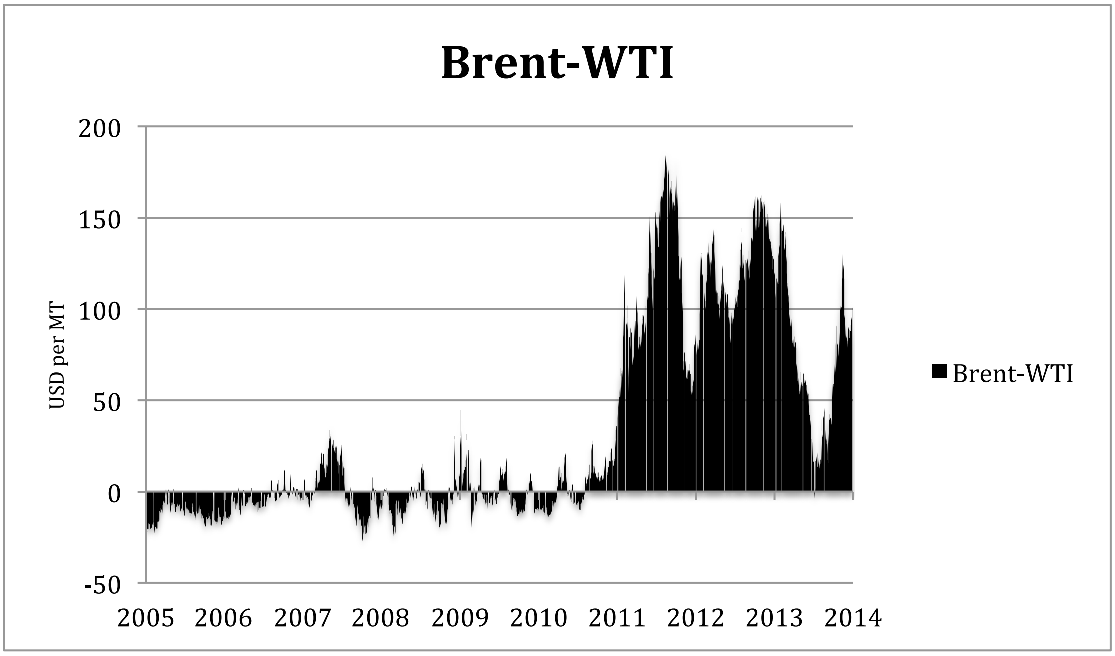

Figure 7 : Brent - WTI spread

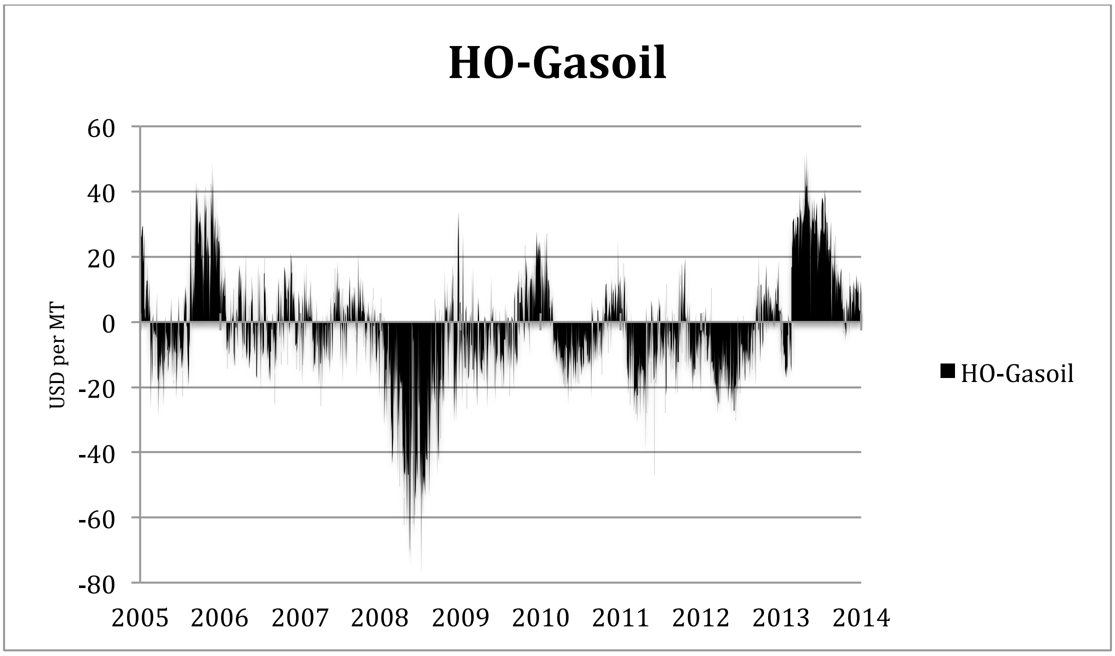

Figure 8 : Heating Oil - Gasoil spread

4.3 Descriptive Statistics

The Minimal Spanning Trees and Hierarchical Trees are constructed from a set of data containing the logarithmic returns of the respective commodities. Define the logarithmic return as:

$$r_t = \ln(X_t - X_{t-1}) \tag{4}$$

Where X_t is the closing price of the commodity X at the time t.

Basic descriptive statistics for the daily logarithmic returns for the whole, pre-crises, crises and post-crises periods are presented in Table 2, Table 3, Table 4 and Table 5 respectively.

We note that an elevated volatility during the crises period. It is higher than both that of the pre- and post-crises periods. The skewness of the logarithmic returns changed during the period of analysis. Pre-crises, all commodities were positively skewed whereas both during and after the crises they were negatively skewed. The Ex. Kurtosis of the Canola seed increased after the crises.

DESCRIPTIVE STATISTICS FOR THE FULL PERIOD
(13 JAN 2005-13 JAN 2014)
Daily logarithmic returns

Commodity	Mean	Min	Max	St. Dev.	Skewness	Ex. Kurtosis
Crude Oil - Brent	0,0004	-0,1048	0,1288	0,0200	-0,1336	0,7514
Crude Oil – WTI	0,0003	-0,1143	0,1278	0,0216	-0,1113	0,4890
Gasoil	0,0004	-0,0902	0,1073	0,0182	0,0605	-0,6284
Heating Oil	0,0003	-0,0968	0,0991	0,0199	0,0174	-0,7965
Palm Oil	0,0003	-0,1090	0,0953	0,0176	-0,2398	1,0026
Soybean Oil	0,0003	-0,0714	0,0808	0,0157	0,1202	-0,7473
Canola seed	0,0002	-0,1404	0,0644	0,0139	-0,7336	5,6707
Rapeseed	0,0002	-0,0763	0,0522	0,0107	-0,5738	1,9304

Table 2 : Descriptive Statistics for the Full Period

DESCRIPTIVE STATISTICS FOR FOR PRE-CRISES PERIOD (13 JAN 2005-9 JUL 2007) Daily logarithmic returns						
Commodity	**Mean**	**Min**	**Max**	**St. Dev.**	**Skewness**	**Ex. Kurtosis**
Crude Oil - Brent	0,0008	-0,0486	0,0563	0,0168	0,0581	-3,1101
Crude Oil – WTI	0,0007	-0,0488	0,0638	0,0179	0,0474	-3,0026
Gasoil	0,0008	-0,0617	0,0758	0,0186	0,0184	-2,3583
Heating Oil	0,0007	-0,0486	0,0965	0,0205	0,3847	-2,2789
Palm Oil	0,0011	-0,0668	0,0556	0,0118	0,0165	1,5548
Soybean Oil	0,0010	-0,0618	0,0808	0,0146	0,4942	-0,6385
Canola seed	0,0007	-0,0498	0,0605	0,0120	0,7877	0,5875
Rapeseed	0,0006	-0,0247	0,0522	0,0073	0,6510	1,8643

Table 3 : Descriptive Statistics for pre-crises period

DESCRIPTIVE STATISTICS FOR CRISES PERIOD (10 JUL 2007-26 JAN 2010) Daily logarithmic returns						
Commodity	**Mean**	**Min**	**Max**	**St. Dev.**	**Skewness**	**Ex. Kurtosis**
Crude Oil - Brent	-0,0001	-0,1048	0,1288	0,0280	-0,0645	-0,8697
Crude Oil – WTI	0,0001	-0,1143	0,1278	0,0300	-0,0777	-1,0095
Gasoil	-0,0001	-0,0902	0,1073	0,0238	0,1737	-1,3890
Heating Oil	0,0005	-0,0486	0,0553	0,0183	-0,0098	-3,3093
Palm Oil	-0,0001	-0,1090	0,0953	0,0258	-0,1496	-1,6602
Soybean Oil	0,0000	-0,0714	0,0744	0,0206	-0,0069	-1,9339
Canola seed	-0,0001	-0,0822	0,0644	0,0178	-0,5590	0,6786
Rapeseed	-0,0001	-0,0573	0,0495	0,0138	-0,3315	-1,1496

Table 4 : Descriptive Statistics for the crises period

DESCRIPTIVE STATISTICS FOR POST-CRISES PERIOD (27 JAN 2010-13 JAN 2014) Daily logarithmic returns						
Commodity	**Mean**	**Min**	**Max**	**St. Dev.**	**Skewness**	**Ex. Kurtosis**
Crude Oil - Brent	0,0004	-0,0898	0,0679	0,0150	-0,3744	-0,3822
Crude Oil – WTI	0,0002	-0,0895	0,0890	0,0168	-0,2374	-0,5045
Gasoil	0,0004	-0,0620	0,0474	0,0131	-0,2356	-1,8282
Heating Oil	0,0004	-0,0846	0,0619	0,0143	-0,2429	-0,9362
Palm Oil	0,0000	-0,0951	0,0508	0,0137	-0,3608	-0,3698
Soybean Oil	0,0000	-0,0486	0,0551	0,0123	0,1076	-1,8597
Canola seed	0,0001	-0,1404	0,0509	0,0121	-1,8243	16,3240
Rapeseed	0,0002	-0,0763	0,0421	0,0103	-1,0327	3,9922

Table 5 : Descriptive Statistics for the post-crises period

4.4 Stationarity tests

The Minimal Spanning and Hierarchical Trees are constructed from the distance matrix $\mathbb{D}$. The distance matrix, in turn, is mapped from the correlation matrix $\mathbb{C}$. As a condition the time-series must be weak-stationary (Mantegna 1999).

Table 6 show the results for the: 1) Augmented Dickey-Fuller Test with a Drift (ADF-ARD), 2) Augmented Dickey-Fuller without drift (ADF-AR) and the 3) Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests for the whole period of all the commodities. All tests indicate that the time-series are all stationary for the whole period. In other words, the crises period did not break the stationarity of the time-series.

STATIONARITY TESTS

Daily logarithmic returns of the full period (13 JAN 2005-13 JAN 2014)

Commodity	ADF-ARD	p-Value	ADF-AR	p-Value	KPSS
Crude Oil - Brent	-50,9373	0,0001	-50,9297	0,0001	0,0534
Crude Oil – WTI	-49,5121	0,0001	-49,5142	0,0001	0,0415
Gasoil	-48,4804	0,0001	-48,4720	0,0001	0,0685
Heating Oil	-48,4826	0,0001	-48,4787	0,0001	0,0542
Palm Oil	-49,0359	0,0001	-49,0341	0,0001	0,0528
Soybean Oil	-46,9335	0,0001	-46,9277	0,0001	0,0628
Canola seed	-44,8082	0,0001	-44,8077	0,0001	0,0728
Rapeseed	-43,6577	0,0001	-43,6466	0,0001	0,0976

Table 6 : Stationarity Tests

5 Results

In this section, we present the Minimal Spanning and Hierarchical Trees associated with our data. The results are divided in four parts, each representing one of the periods analysed. Each section, in turn, is divided in three sub-sections that represent the daily, weekly and monthly returns. In the graphs, the distance d_{ij} is presented in regular font and the bootleg value for the link is presented inside brackets.

5.1 Whole Sample (13 JAN 2005-13 JAN 2014)

Figure 9, Figure 10 and Figure 11 present the MST and HT for the daily, weekly and monthly logarithmic returns respectively. For all frequencies there are two clusters. One consisting of the energy commodities *Brent*, *WTI*, *HO* and *Gasoil* and one of the agricultural commodities *Canola*, *CPO*, *Rapeseed* and *SBO*. We will refer to them as the Energy and the Agriculture cluster respectively. The MSTs for all the frequencies infer that the two clusters are linked via *SBO-Brent*. This distance connecting the two clusters is the second longest for the daily returns and the longest for both the weekly and monthly returns. The same is true for the bootstrap values. This indicates that the two clusters are not that well connected and that the interconnection between the energies and agricultural commodities is not that prominent when analysing the whole period.

An analysis of the links within the energy cluster indicates that there are two sub-clusters, one for the crude oils: *WTI* and *Brent* and one for the fuels: *Gasoil* and *Heating Oil*. The connections between these two sub-clusters are from *Brent* to *HO* for all frequencies. One possible explanation is that *Brent* is now the *de facto* crude oil index of the world and that *Heating Oil* is a more liquid market than *Gasoil*.

Analysing the agricultural cluster points out that the distances between the nodes are greater than those for the energy cluster. On top of this, the bootstrap values are lower as well. As such, the cluster is looser for all frequencies. In particular, the connection for *CPO* with the other commodities changes when changing the return frequency. For the daily returns, *CPO* is connected to *Rapeseed* by the greatest distance in the graph but its bootstrap value is high. The MSTs and HTs for the weekly and monthly returns connect *CPO* directly to *SBO*. Intuitively, the link between *CPO* and *SBO* should be strong as both are vegetable oils and are, to some extent, interchangeable. That this is not the case for the daily returns may be explained by the fact that *BMD* (where *CPO* trades) is closed when *CBOT* (where *SBO* trades) is liquid. The markets in the United States are the most developed and can be seen to drive the prices across the globe. Consequently, *CPO* is expected to react to *SBO*'s price changes only the day afterwards. However, the effect of this lag should diminish when we change frequencies from daily to weekly and monthly.

By analysing the average distances and average bootstrap values we can deduce the reliability of the entire graph. The results in Table 7 unanimously show that the graph's reliability increases as we go from daily to weekly and monthly returns. Most importantly the link between the clusters is at its strongest for the monthly returns (see Figure 11).

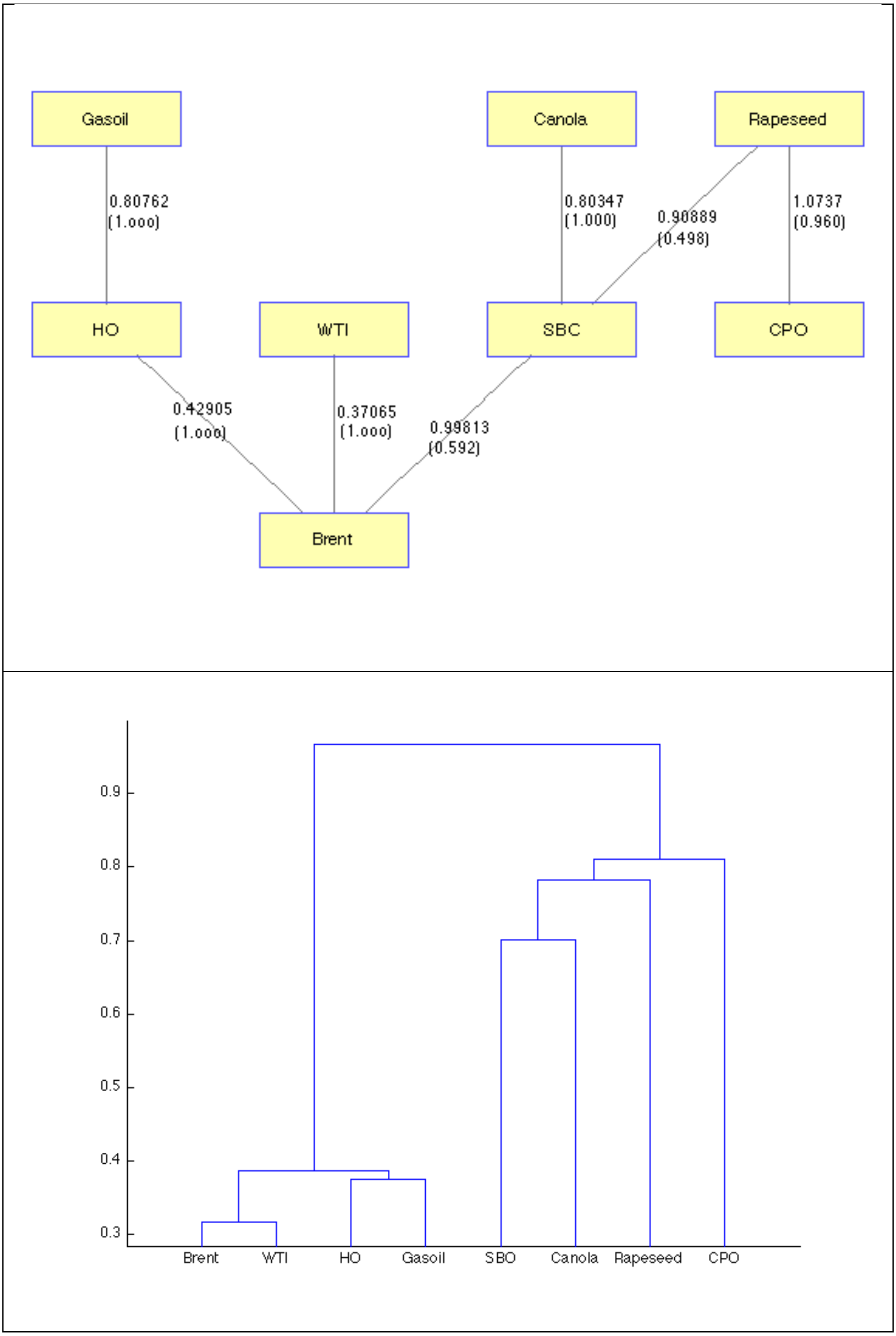

Figure 9 : MST and HT of Daily Returns for the Full Period

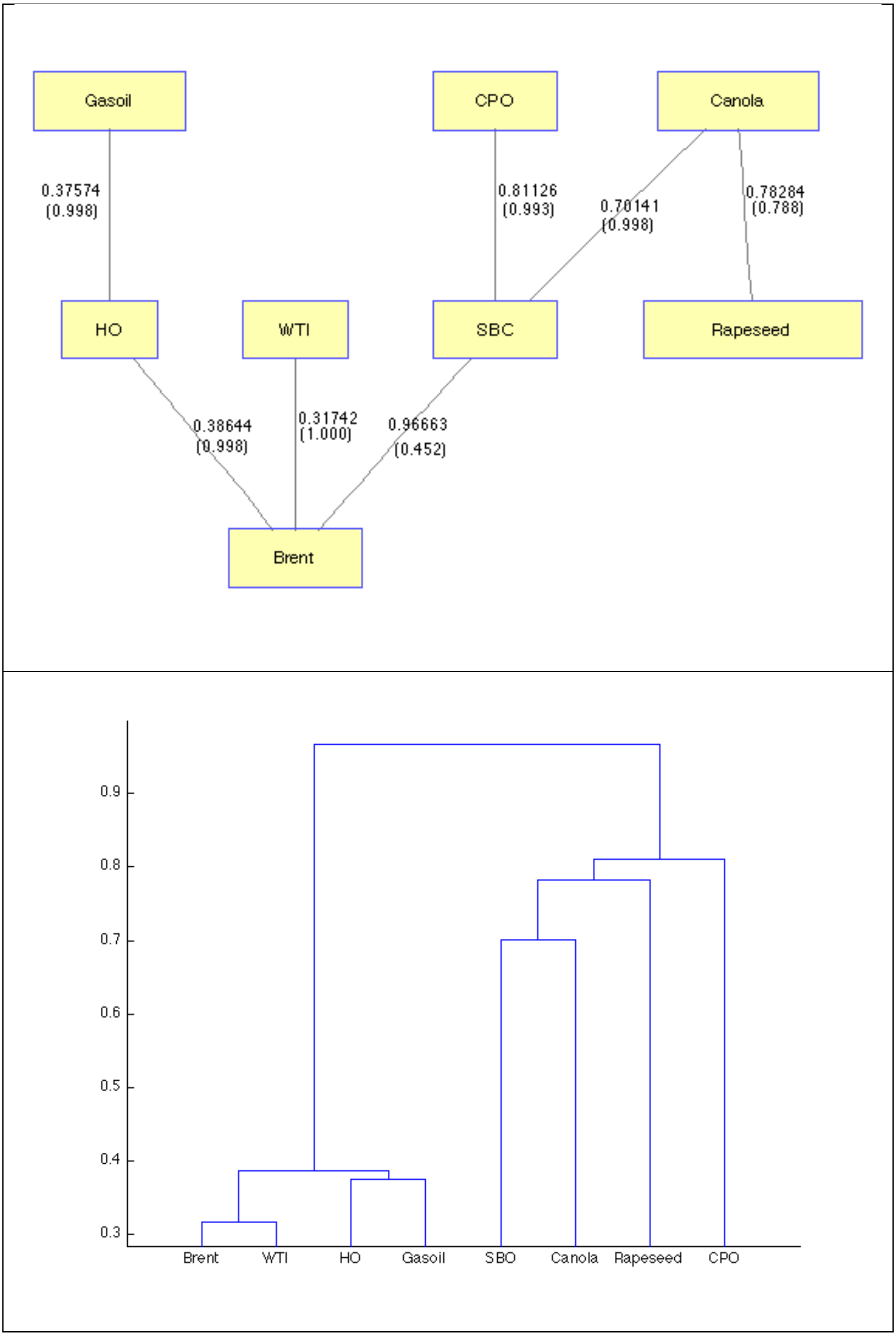

Figure 10 : MST and HT of Weekly Returns for the Full Period

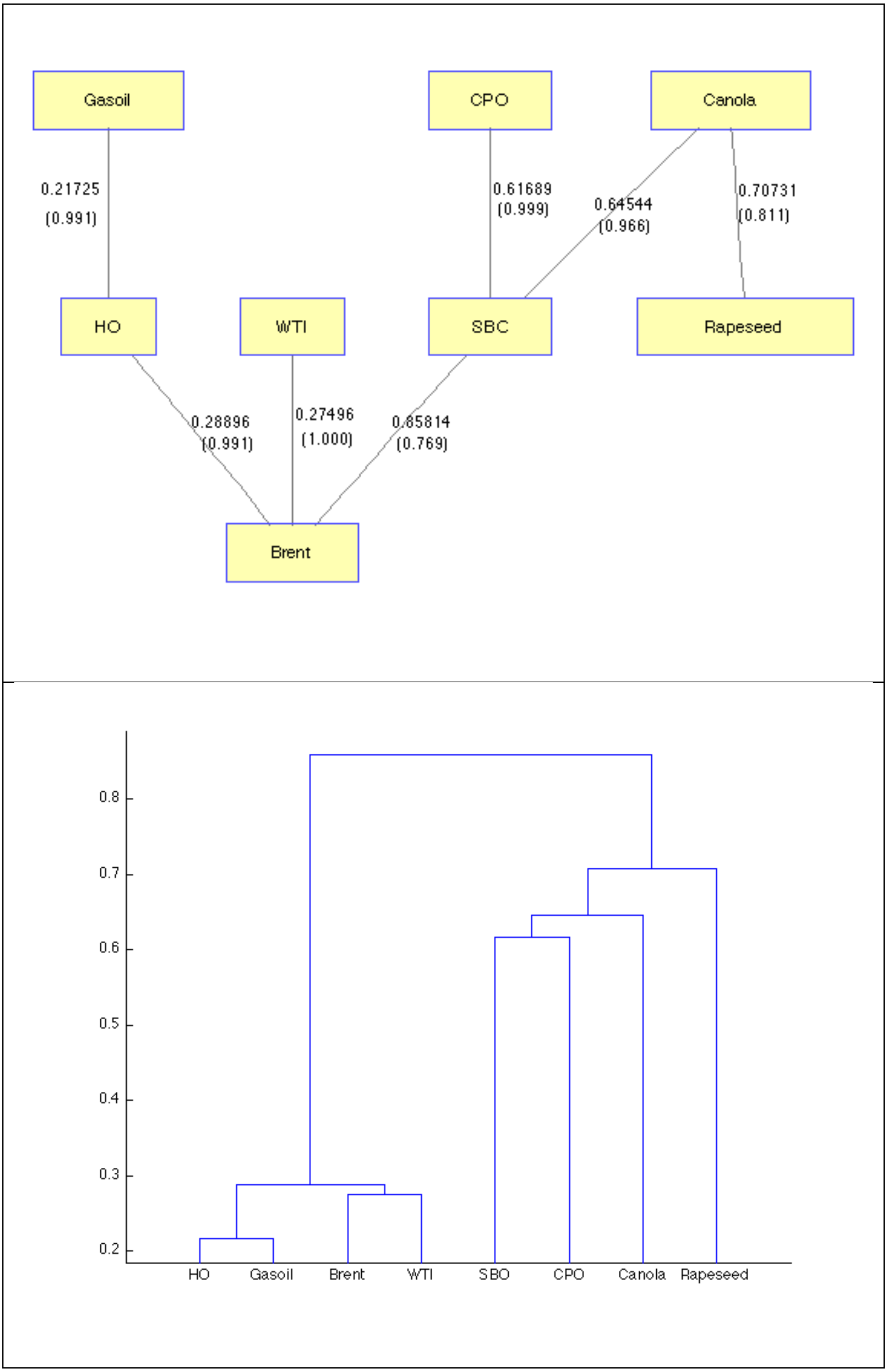

Figure 11 : MST and HT of Monthly Returns for the Full Period

Period	Frequency	$\overline{d_{ij}}$	$\overline{b_{ij}}$
Whole period	Daily	0.7702	0.8643
	Weekly	0.6202	0.8896
	Monthly	0.5156	0.9324

Table 7 : Average Distances and Bootstrap Values for the Whole Period

5.2 Pre-crises Period (13 JAN 2005-9 JUL 2007)

The MSTs and HTs for the pre-crises period differ in many ways from those of the whole period. One pattern that remains intact is the division of the commodities in two clusters: energy and agriculture. Whereas the two clusters were linked via the *SBO-Brent* vertex for all frequencies when analyzing the full period they now change as the frequencies change. *SBO-Brent*, *SBO-Gasoil*, *SBO-WTI* connect the clusters for the daily, weekly and monthly returns respectively. The connecting vertices are all longer during the pre-crises period than those for the whole sample. Furthermore, their bootstrap values are low. This combination indicates that the clusters were not as closely interconnected prior to the crises period as they were during the whole period.

Addressing the energy cluster, the same two sub-clusters of *Brent-WTI* and *Gasoil-HO* are present in the weekly and monthly HTs. For all the frequencies, the distances within the energy sub-cluster were greater during the pre-crises period than for the whole period. Further indicating that the commodities were less interconnected prior to the crises.

The looseness was present in the agricultural cluster as well. This is an indication that the agricultural commodities behaved more independently prior to the crises.

Within the agricultural cluster, *CPO* shows the same behaviour as for the full period. Again, the market hours may be one of the causes. *Rapeseed* proved to be quite loosely related to the rest in the group prior to the crises as well.

The average distances were greater for the pre-crises period than for the whole sample. Interestingly, the bootstrap values were higher too which indicates that their strength was greater even though the nodes were further from each other. This supports the theory that the commodities were less interrelated prior to the crises than they were during and after it.

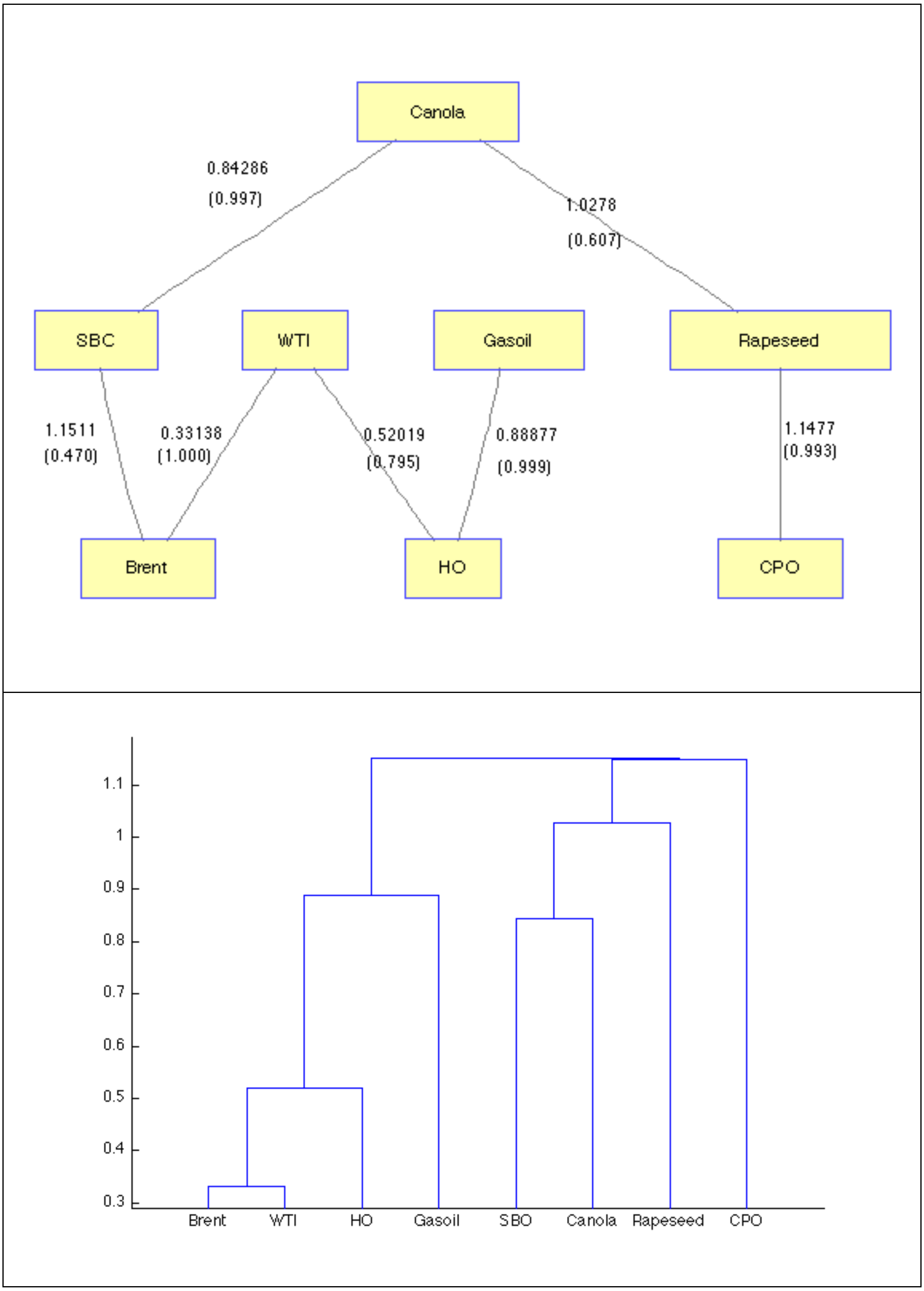

Figure 12 : MST and HT of Daily Returns for the Pre-crises Period

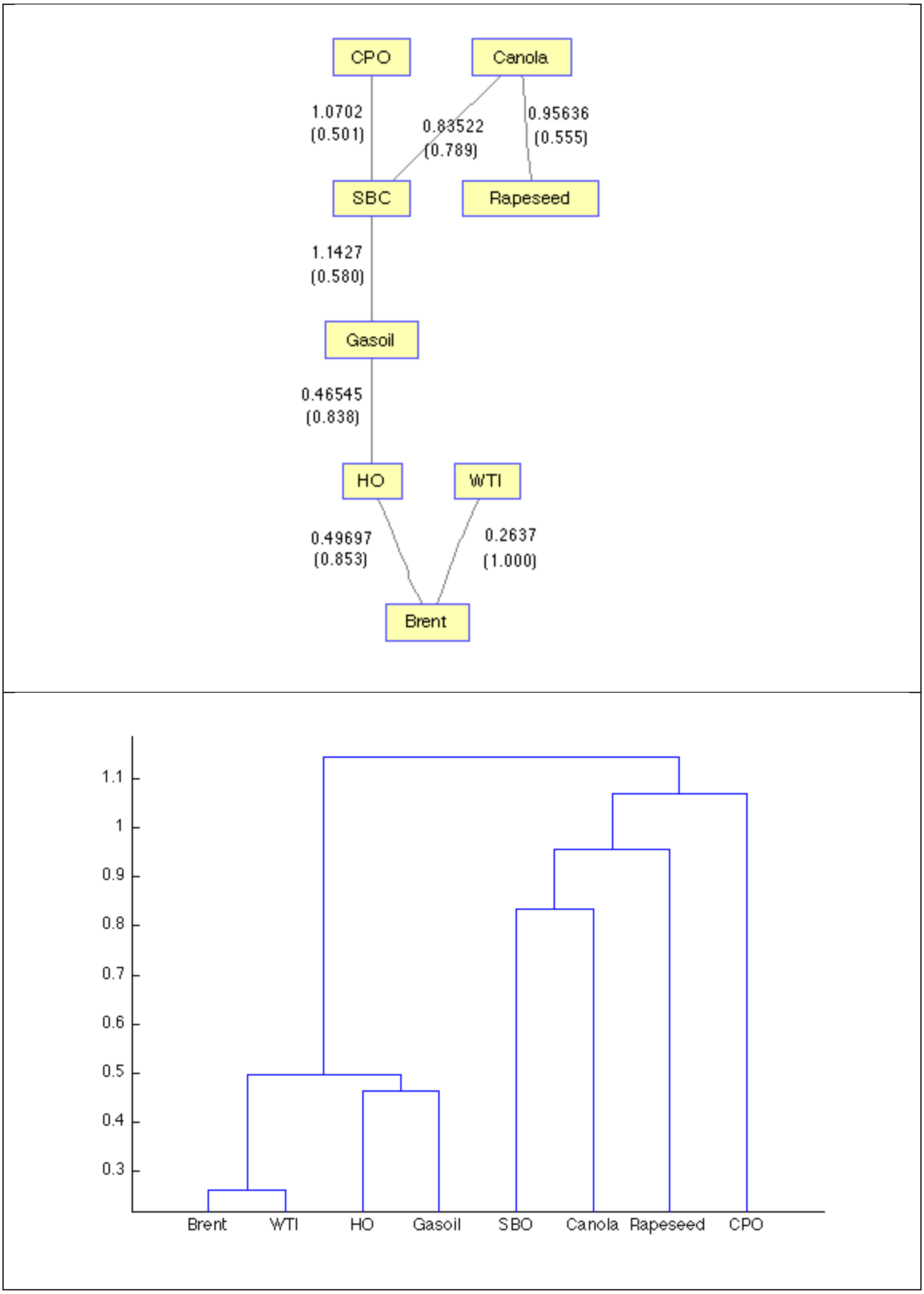

Figure 13 : MST and HT of Weekly Returns for the Pre-crises Period

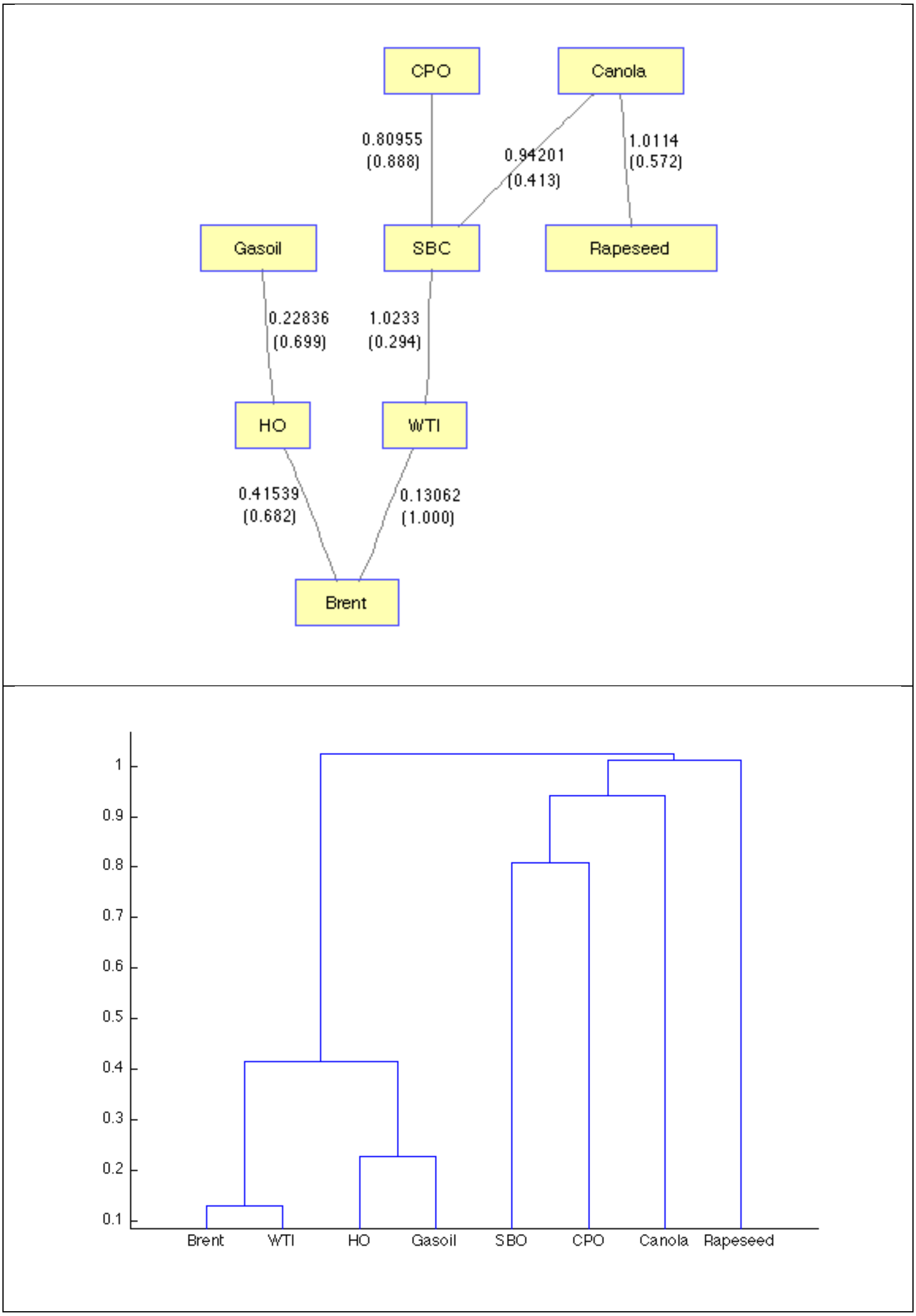

Figure 14 : MST and HT of Monthly Returns for the Pre-crises Period

Period	Frequency	$\overline{d_{ij}}$	$\overline{b_{ij}}$
Pre-crises	Daily	0.8443	0.8283
	Weekly	0.6044	0.7309
	Monthly	0.6515	0.6497

Table 8 : Average Distances and Bootstrap Values for the Pre-crises Period

5.3 Crises Period (10 JUL 2007-26 JAN 2010)

The MSTs and HTs for the crises period form the same two clusters as we have seen before: energy and agricultural. The two clusters are linked via *SBO-Brent*, *SBO-HO* and *SBO-Brent* for the daily, weekly and monthly returns respectively. The links that connect the two clusters are on average shorter and stronger than those for the whole period and in particular than those for the pre-crises period. This indicates that the various commodities showed greater co-movement during the crises than they did before.

Looking at the energy cluster, we see the same two sub-clusters of crude oil and fuel as we have seen before. The HTs for the pre-crises and crises periods are similar for the energies.

The agricultural sub-clusters are denser than they were prior to the crises for all frequencies. The bootstrap values are on average slightly higher and in particular the link *CPO-SBO* is stronger.

For all frequencies, the average distances are shorter than those for the whole period as well as pre-crises. However, the bootstrap values are on average only better than those for the pre-crises period. This may stem from the smaller sample size.

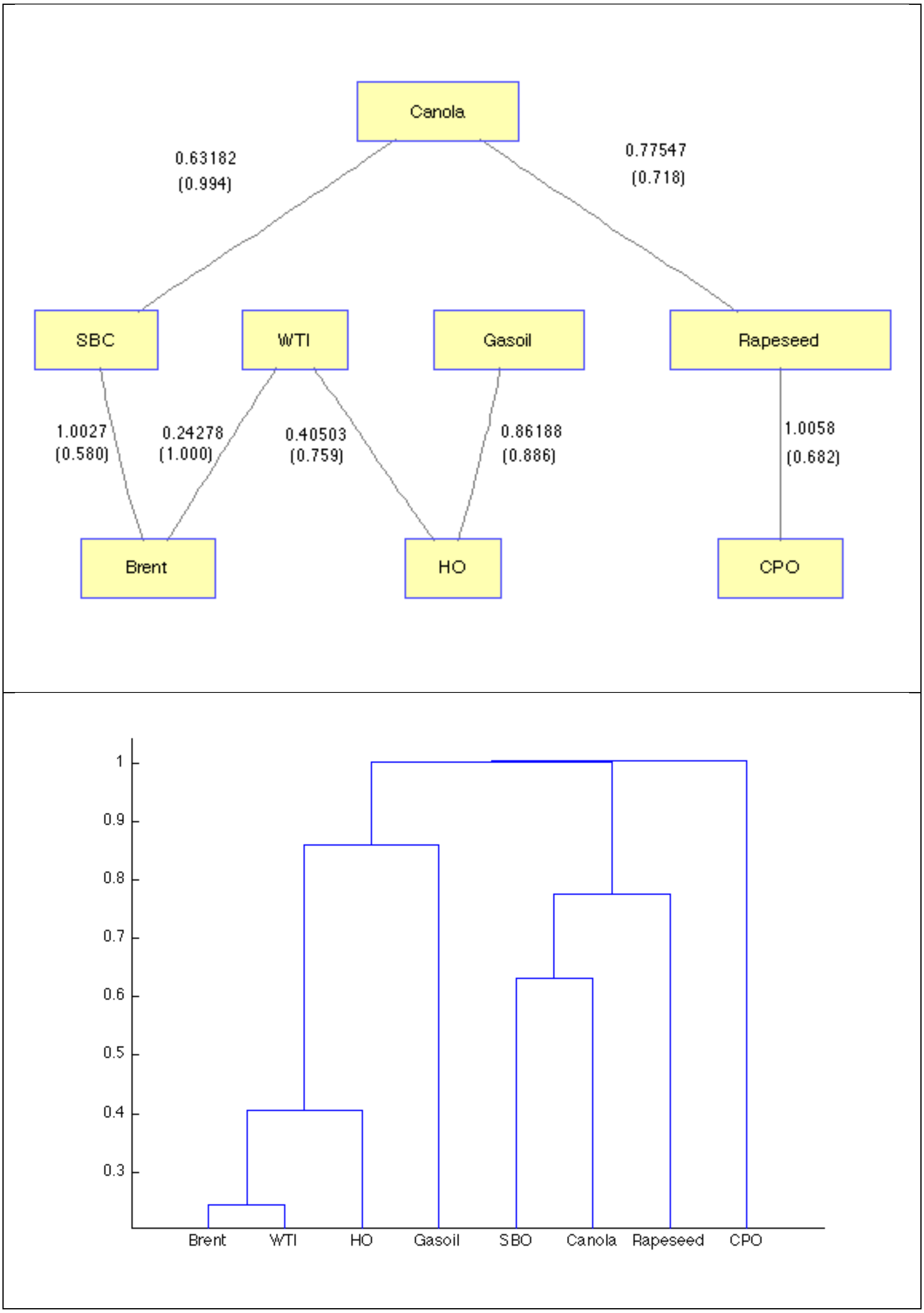

Figure 15 : MST and HT of Daily Returns for the Crises Period

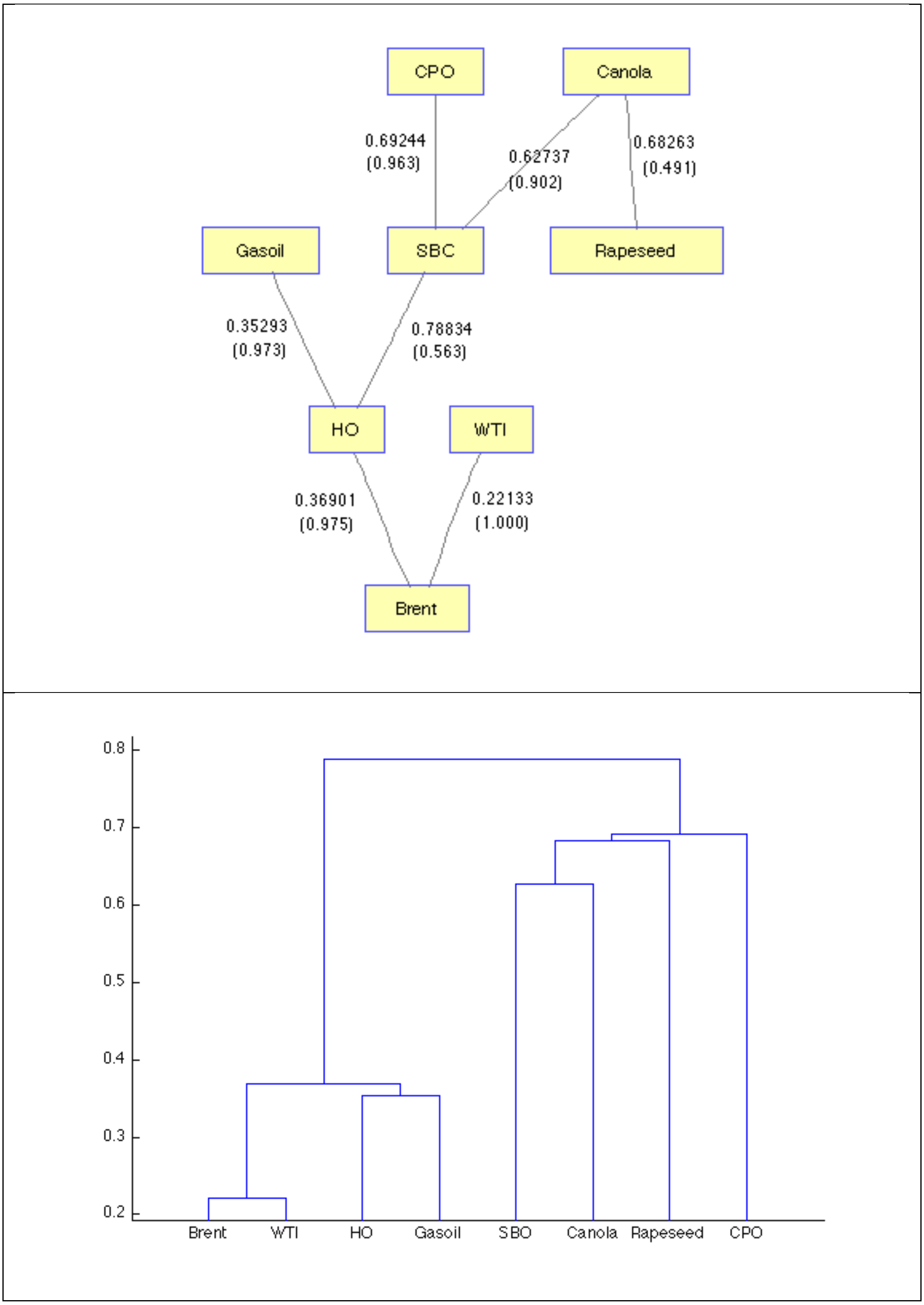

Figure 16 : MST and HT of Weekly Returns for the Crises Period

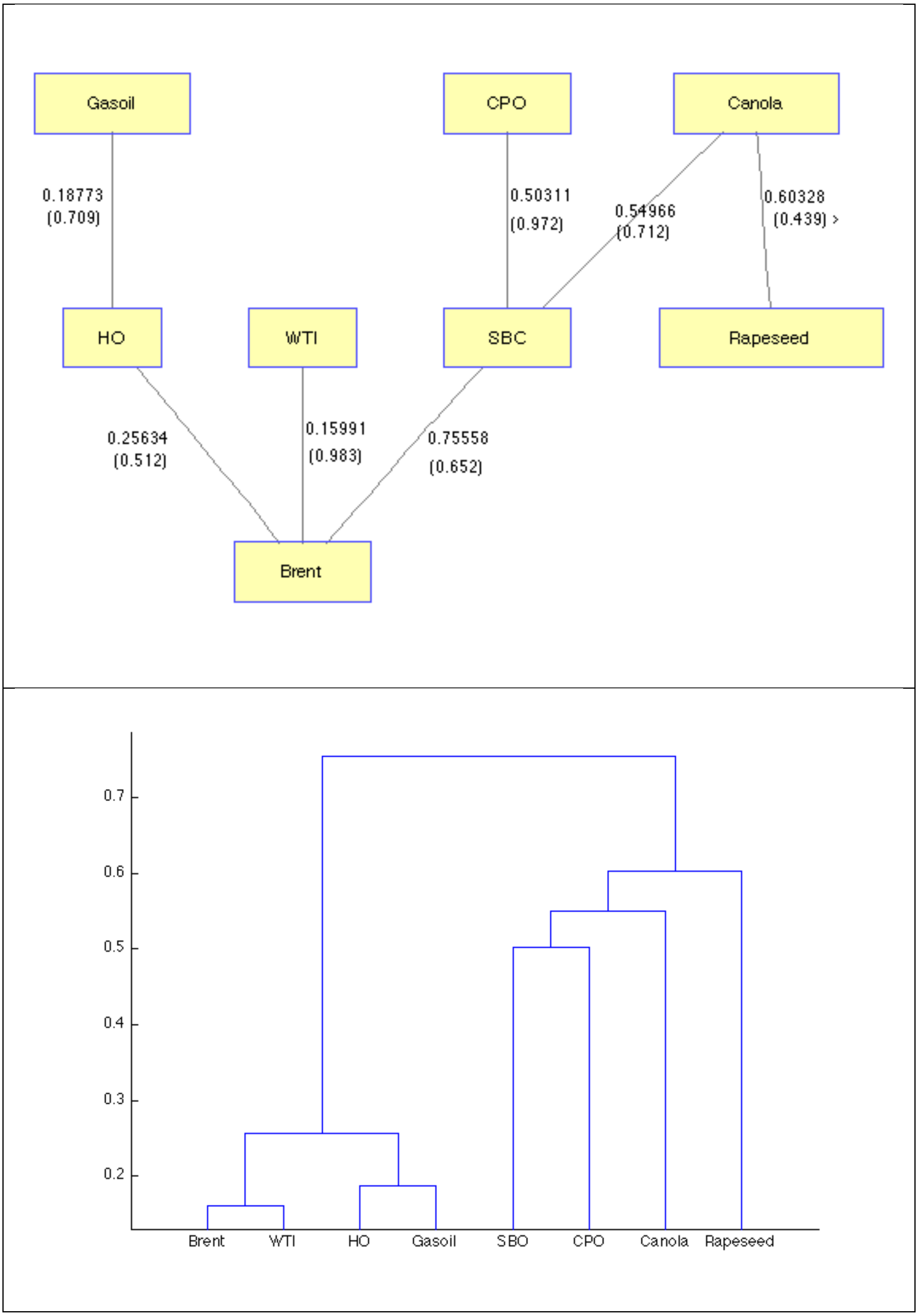

Figure 17 : MST and HT of Monthly Returns for the Crises Period

Period	Frequency	$\overline{d_{ij}}$	$\overline{b_{ij}}$
Crises	Daily	0.7011	0.8027
	Weekly	0.5334	0.8381
	Monthly	0.4308	0.7113

Table 9 : Average Distances and Bootstrap Values for the Crises Period

5.4 Post-crises Period (27 JAN 2010-13 JAN 2014)

For the post-crises, the MSTs and HTs form the same energy and agricultural clusters as for the other periods. *SBO-Brent*, *SBO-Gasoil* and *SBO-HO* link the two clusters for the respective frequencies. The lengths of these vertices connecting the two clusters are longer than for both the whole and crises period. They are, however, shorter than those prior to the crises. The bootstrap values indicate that the links are stronger than they were before and during the crises but less so than for the whole sample. Which may be a consequence of the size of the sub-samples.

Within the energy cluster one thing has changed. *WTI* is on its own branch for all the MSTs and it is also separated in the HTs. This corresponds with the decoupling of the *WTI-Brent* spread that we have seen due to the shale gas revolution in the United States.

The structure of the agricultural cluster in the MST of daily returns has morphed. *SBO* is the closest connection to all the others. *SBO* is the most liquid of all the agricultural commodities analysed and may react to/drive the other commodities.

The distances within the agricultural cluster are shorter than prior to the crises but longer than during it. The agricultural links' bootstrap values are higher than both prior and during the crises. As such the graphs' reliability are higher than for any other individual period analysed.

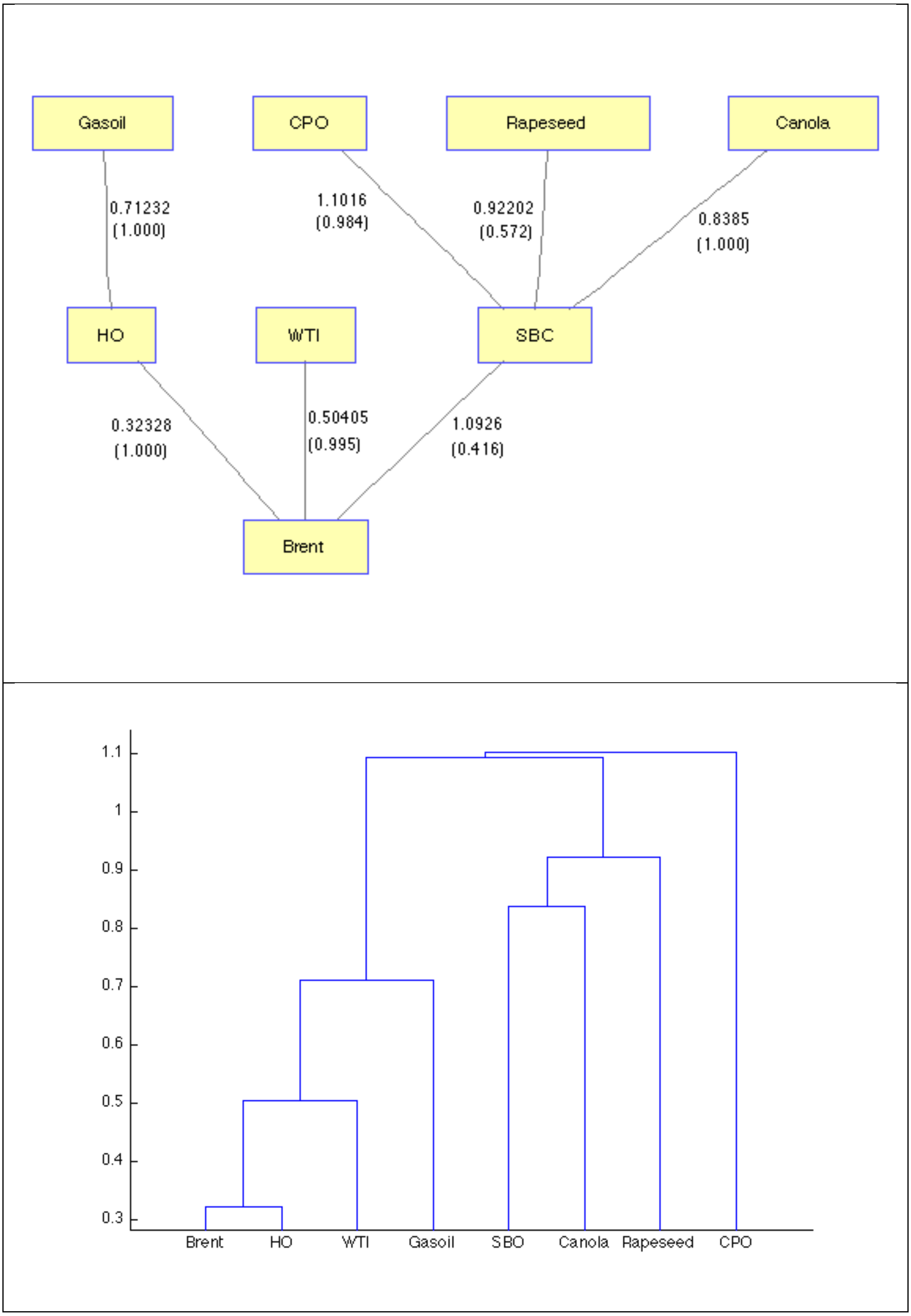

Figure 18 : MST and HT of Daily Returns for the Post-crises Period

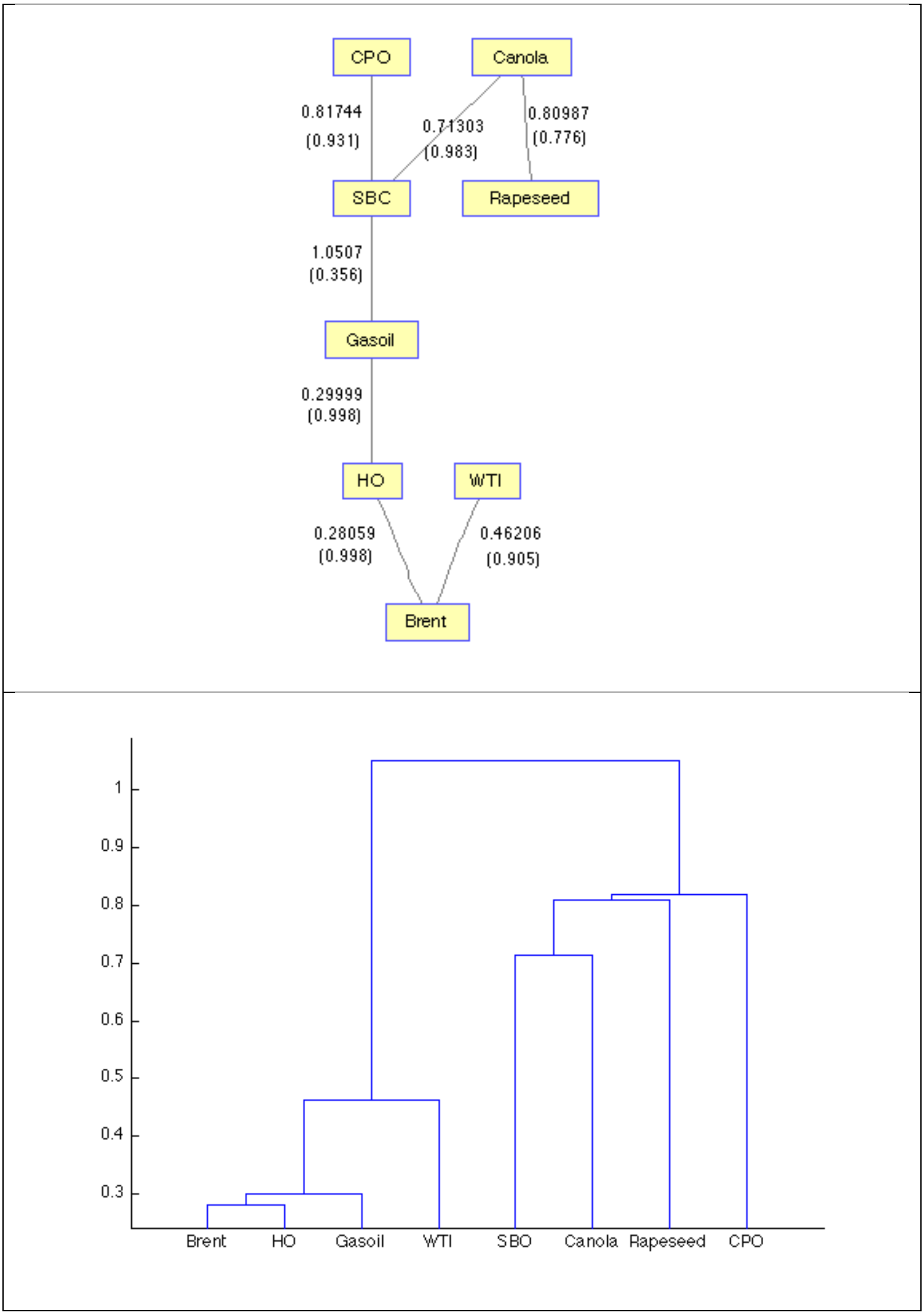

Figure 19 : MST and HT of Weekly Returns for the Post-crises Period

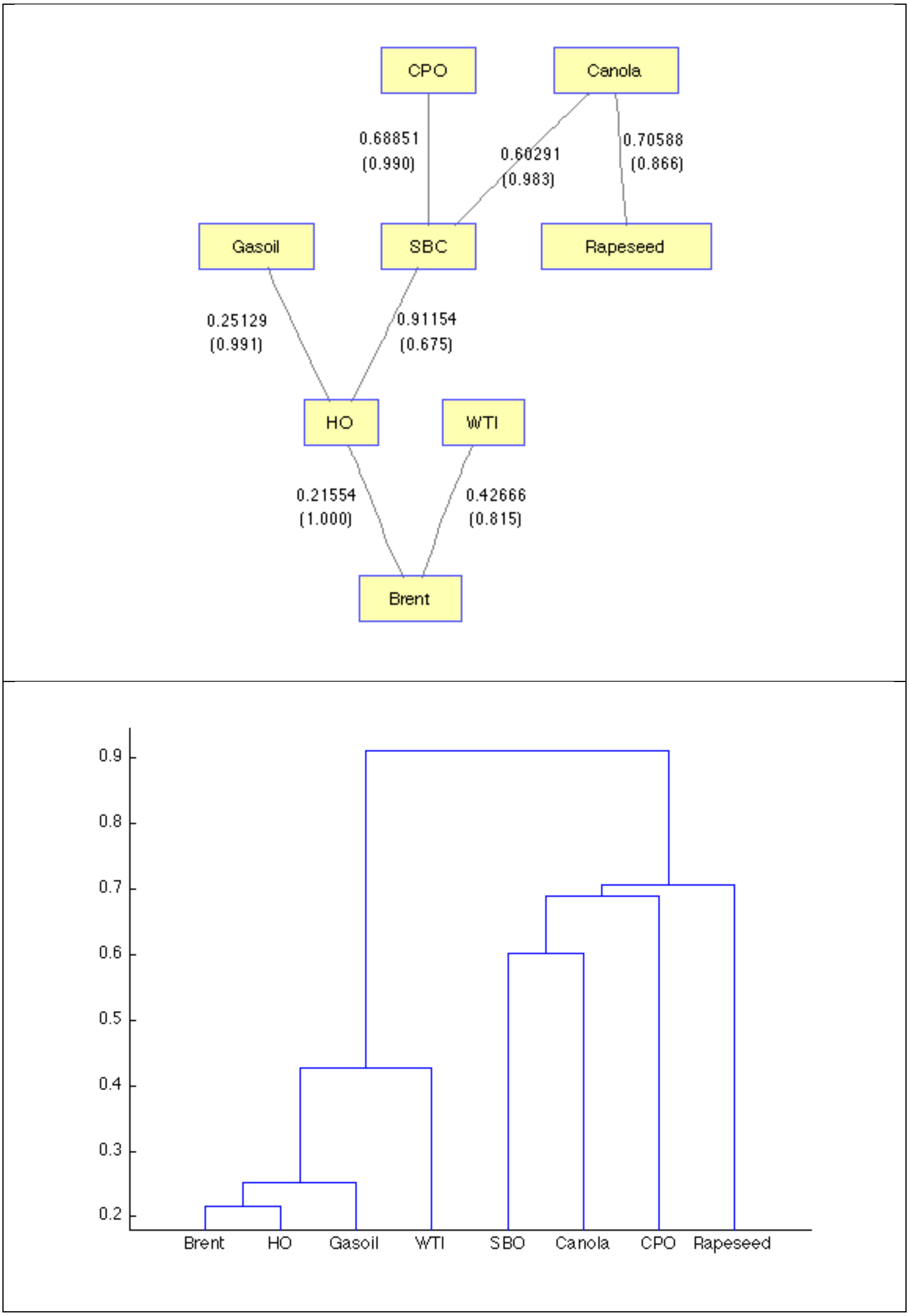

Figure 20 : MST and HT of Monthly Returns for the Post-crises Period

Period	Frequency	$\overline{d_{ij}}$	$\overline{b_{ij}}$
Post-crises	Daily	0.7849	0.8524
	Weekly	0.6334	0.8496
	Monthly	0.5432	0.9029

Table 10 : Average Distance and Bootstrap Value for the Post-crises Period

6 Conclusion

The aim of this thesis was to investigate whether the substitution of illiquid and non-tradable indices/commodities with their closest exchange traded alternative would be a fair model to represent the price transmission network in the biodiesel space.

During this process, we found that the commodities formed two sub-networks: one consisting of energy and the other of agricultural commodities. The two clusters were always linked together via soybean oil from the agricultural sub-network whereas the connecting energy commodity varied depending on period and data frequency. Furthermore, the length and strength of the node connecting the two clusters was also dependent on the frequency and data period. In general, the node connecting the two clusters as well as the rest of the network's nodes strengthened as we moved from daily to weekly and to monthly price data. This indicates that shocks may distort the spreads between commodities in the short run but that their intra-relationship should revert to that previously seen unless there has been a structural change.

Such structural changes can for example be caused by macro-economical events such as the food and financial crises or by regulatory change such as new biofuels legislation. During our sample period, there were examples of both and we saw that the length of the connecting vertex shrunk during the crises period but it did not reach its previous length once the crises were over. Not only the connecting vertex, but also the rest of the network got denser and stronger during the crises period and the network has so far remained denser than prior to the crises.

One possible explanation for the commodities' increased co-movement during the crises period may be that many investors and commercials were forced to limit their risk due to the extreme market conditions and this may have made them close all of their positions no matter the commodity.

That the network has remained denser than prior to the crises may be an effect of a continued risk averseness where commercials and other participants have started to trade premium/basis instead of flat outright. Increased premium/basis trading may also explain why we found soybean oil to be the centre of the agricultural sub-network after the crises period as some may have started to price other commodities over the one that is the most liquid in order to be sure that they will be able to find a counter-part in the most difficult of markets.

In summary, the data does not confirm the hypothesis that the substitution of cash markets and indices with tradable cleared futures would be a fair representation of the intra-commodity links in the biodiesel space. We expected rapeseed and/or canola as the substitute for rapeseed/canola oil to have a much more prominent role. Instead, soybean oil, which is not as commonly used for biodiesel, proved to be the link between the two clusters.

To improve upon the results in the future there are a number factors that should be addressed. First of all, the usage of commodities that trade in different time zones made a small data loss necessary and time differences inevitable. Secondly, the fact that the commodities trade on different market places and in different currencies made them susceptible to foreign exchange risk, which in the case of *e.g.* Malaysia where the Palm Oil is traded may be severe. Thirdly, the geographical location of the market places only reflects the fact that the various commodities are found in different locations around the world that may be different from their main consumption areas. Consequently, the price of the freight should be considered when comparing FOB Malaysia with FOB Europe. Finally, substituting rapeseed and canola oil with their respective seed may have had disproportionate effects that weakened the link between energy and rapeseed/canola oil.

For future researchers it may be possible to circumvent these problems of time differences, foreign exchange, freight spreads and substitution of commodities by using data from the cash market in Europe. It would be interesting to see how the network formed if CIF Rotterdam Crude Palm Oil, Ex-Dutch Mill Rapeseed Oil and Ex-Dutch Mill Soybean Oil prices were used. ICE Gasoil would still represent the energy complex. However, by acknowledging the need to the trade the cash market only reaffirms our rebuttal of the original hypothesis that it is possible to keep a price transmission network of physical commodities intact by replacing cash markets with futures.

7 Bibliography

United States Department of the Treasury. "Troubled Asset Relief Program : Two Year Retrospective." Office of Financial Stability, United States Department of the Treasury, 2010.

Zilberman, David, Gal Hochman, Deepak Rajagopal, Steve Sexton, and Govinda Timilsina. "The Impact of Biofuels on Commodity Food Prices: Assessment of Findings." *American Journal of Agricultural Economics* 95, no. 2 (2012): 275-281.

Zhang, Z, L Lohr, and M Wetzstein. "Mitigating volatile U.S. gasoline prices and itnernalizing external costs : A win-win fuel portfolio." *American Journal of Agricultural Economics*, no. 90 (2008): 1218-1225.

Zhang, Z, L Lohr, C Escalante, and M Wetzstein. "Ethanol, corn, and soybean price relations in a volatile vehicle-fuels market." *Energies*, no. 2 (2009): 320-339.

Wang, Gang-Jin, Chi Xie, Feng Han, and Bo Sun. "Similarity measure and topology evolution of foreign exchange markets using dynamic time warping method: Evidence from minimal spanning tree." *Physica A: Statistical Mechanics and its Applications* 391, no. 16 (2012): 4136-4146.

Yergin, Daniel. *The Prize: The Epic Quest for Oil, Money & Power.* New York: Free Press, 2008.

Bonanno, G, G Caldarelli, F Lillo, S Micciche, N Vandewalle, and R Mantegna. "Networks of equities in financial markets." *European Physical Journal B*, no. 38 (2004): 363-371.

European Biodiesel Board. *http://www.ebb-eu.org/stats.php#.* 18 04 2014. http://www.ebb-eu.org/stats.php# (accessed 04 18, 2014).

Efron, B. "Bootstrap Methods: Another Look at the Jacknife." *The Annals of Statistics* 7, no. 1 (January 1979): 1-26.

Du, X, C.L. Yu, and D.J. Hayes. "Speculation and volatility spillover in the crude oil and agricultural commodity markets : A Bayesian analysis." *Energy Economics*, no. 33 (2011): 497-503.

Felsenstein, Joseph. "Phylogenies and the Comparative Method." *The American Naturalist* 125, no. 1 (January 1985): 1-15.

Jung, Anna, Philipp Dörrenberg, Anna Rauch, and Michael Thöne. *Government support for ethanol and biodiesel in the European Union – 2010 Update.* Geneva: For the Global Subsidies Initiative (GSI) of the International Institute for Sustainable Development (IISD) Geneva, Switzerland, 2010.

Janda, Karel, Ladislav Kristoufek, and David Zilberman. "Bioduels: policies and impacts." *Agric. Econ. - Czech* 8, no. 58 (2012): 372-386.

Janda, Karel, Ladislav Kristoufek, and David Zilberman. "Biofuels: policies and impacts." *Agricultural Economics - Czech* 58, no. 8 (2012): 372-386.

Knothe, Gerhard. "Historical perspectives on vegetable oil-based diesel fuels." *Inform* 12 (November 2011): 1103-1107.

Kruskal, Joseph B. “On the shortest spanning subtree of a graph and the travelling salesman problem.” *Proceedings of the American Mathematical Society* 7, no. 1 (1956): 48-50.
Kristoufek, Ladislav, Karel Janda, and David Zilberman. “Correlations Between Biofuels and Related Commodities Before and During the Food Crisis: A Taxonomy Persepective.” Working Paper, UC Center for Energy and Environmental Economics, Berkeley, 2012.
Kristoufek, Ladislav, Karel Janda, and David Zilberman. *Non-Linear Price Transmission Between Biofuels, Fuels and Food Commodities.* Prague, February 2013.
Mantegna, R N. “Hierarchical structure in financial markets.” *The European Physical Journal B*, no. 11 (1999): 193-197.
OECD. *Biofuel Support Policies: an Economic Assessment.* Pari: OECD, 2008.
Pelkmans, Luc, Leen Govaerts, Panagiotis Georgopoulos, Villy Portouli, and Tobias Wiesenthal. “ASSESSMENT OF MEASURES TO SUPPORT BIOFUEL INTRODUCTION IN EUROPEAN COUNTRIES.” *PREMIA.* PREMIA, 2007.
Peri, Massimo, and Luca Baldi. *Biodiesel and vegetable oil market in European Union: some evidences from threshold cointegration analysis.* EAAE, 2008.
Peri, Massimo, and Lucia Baldi. “The effect of biofuel policies on feedstock market: Empirical evidence for rapeseed oil prices in EU.” *Reource and Energy Economics* (Elsevier), no. 35 (2013): 18-37.
Serra, T, D Zilberman, and J.M. Gil. “Price volatility in ethanol markets.” *European review of Agricultural Economics*, no. 38 (2011): 259-280.
Serra, T, D Zilberman, J.M. Gil, and B.K. Goodwin. “Nonlinearities in the U.S. corn-ethanol-oil-gasoline price system.” *Agricultural Economics*, no. 42 (2011): 35-45.
Tumminello, M, T Di Matteo, T Aste, and R N Mantegna. “Correlation based networks of equity returns sampled at different time horizons.” *European Physical Journal B*, no. 55 (2007): 209-217.
Tumminello, Michele, Claudia Coronnello, Fabrizio Lillo, Salvatore Miccichè, and N Rosario Mantegna. “Spanning Trees and bootstrap reliability estimation in correlation based networks.” February 2008.
Tabak, Benjamin, Thiago Serra, and Daniel Cajueiro. “The expectation hypothesis of interest rates and network theory: The case of Brazil.” *Physica A: Statistical Mechanics and its Applications* 388, no. 7 (2009): 1137-1149.
Tabak, Benjamin, Thiago Serra, and Daniel Cajueiro. “Topological properties of commodities networks.” *European Physical Journal B* 74 (2010): 209-217.
The IntercontinentalExchange. *https://www.theice.com/publicdocs/ICE_Canola_Brochure.pdf.* 08 02 2014. https://www.theice.com/publicdocs/ICE_Canola_Brochure.pdf (accessed 02 08, 2014).
Trujillo-Barrera, Andres, Mindy Mallory, and Philip Garcia. “Volatility Spillovers in the U.S. Crude Oil, Ethanol, and Corn Futures Markets.” *Journal of Agricultural and Resource Economics* 37, no. 2 (2012): 247-262.

Appendix A MATLAB Code

The code used in this thesis.

Appendix A.1 Minimal Spanning Tree – minSpanTree.m

```
A = input('Please enter an array > ');
temp_labels = input('Please enter an array with the node names inside { }
>');

temp_CorrMat = corr(A);
temp_DistMat = tril(sqrt(2*(1-temp_CorrMat)));
temp_UG=sparse(temp_DistMat);
[temp_ST,pred] = graphminspantree(temp_UG,'Method','Kruskal');
view(biograph(temp_UG,temp_labels,'ShowArrows','off','ShowWeights','on'))
view(biograph(temp_ST,temp_labels,'ShowArrows','off','ShowWeights','on'))
```

Appendix A.2 Hierarchical Tree – hierTree.m

```
A = input('Please enter an array > ');
temp_labels = input('Please enter an array with the node names inside { }
>');

B=linkage(squareform(sqrt(2*(1-(corr(A))))),'single');
figure()
dendrogram(B,'Labels',temp_labels)
```

Appendix A.3 Bootstrap Values

Appendix A.3.1 mstBWeights.m

```
BootstrapCorr; %m script that generates the distance matrices

num_myDistMatrices = size(myDist,3);
num_pred = size(myDist,2);
UG = zeros(1,1); %place to store my sparse for the undirected graph
pred = zeros(num_myDistMatrices,num_pred);
%bootstrapCorr = zeros(num_myDistMatrices, num_pred);

for i=1:1:num_myDistMatrices
    UG = sparse(myDist(:,:,i));
    [ST,pred(i,:)] = graphminspantree(UG,'Method','Kruskal');
end

bootstrapCorr = mode(pred);
bootstrapCert = zeros(1,num_pred);

for j=1:1:num_pred
    bootstrapCert(:,j) = sum(pred(:,j)==bootstrapCorr(j));
end


node_num = 1:1:num_pred;

A = [node_num; bootstrapCorr; (bootstrapCert/num_myDistMatrices)];


fprintf('\n');
fprintf('%8s %19s %19s \n','node','primary connection','bootstrap weight');
fprintf('%6.0f %12.0f %23.3f \n',A);
```

Appendix A.3.2 BootstrapCorr.m

```
A = input('Please enter an array > ');
n = input('Please enter number of bootstraps > ');


temp_corrMat = corr(A);
num_rows = size(temp_corrMat,1);
num_cols = size(temp_corrMat,2);

[bootstat,bootsam] = bootstrp(n,'corr',A);
num_cols_bootsam = size(bootsam,2);

myArray = zeros(num_rows,num_cols,num_cols_bootsam);
myDist = zeros(num_rows,num_cols,num_cols_bootsam);

for i=1:1:num_cols_bootsam
    myArray(:,:,i) = tril(corr(A(bootsam(:,i),:)));

    myDist(:,:,i) = tril(sqrt(2*(1-myArray(:,:,i))));
end
```