Australian National University

Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Regime-Dependent Topological Properties of Biofuels Networks

CAMA Working Paper 49/2012
November 2012

Ladislav Kristoufek
Charles University, Prague
Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic

Karel Janda
Charles University, Prague
University of Economics, Prague
CERGE-EI
Centre for Applied Macroeconomic Analysis

David Zilberman
University of California in Berkeley

Abstract

We analyze the relationships between biodiesel, ethanol and related fuels and agricultural commodities with a use of minimal spanning trees and hierarchical trees and Granger causality. We construct this trees for different frequencies (weekly, monthly and quarterly). We find that in short-term, both ethanol and biodiesel are very weakly connected with the other commodities. In medium and long term, the network structure becomes more interesting. We especially concentrate on the links and comovements between the commodities for different stages of the market based on the level of prices of important commodities or groups of commodities. Such approach in general confirms the separation of network into fuels and food branches for majority of analyzed cases.

Keywords

Biofuels; Networks; Causality

JEL Classification

C22; C38; Q42

Suggested Citation:

Kristoufek, L., K. Janda and D. Zilberman (2012). "Regime-Dependent Topological Properties of Biofuels Networks", CAMA Working Paper 49/2012.

Address for correspondence:

(E) cama.admin@anu.edu.au

Regime-dependent topological properties of biofuels networks *

Ladislav Kristoufek[a,b], Karel Janda[a,c,d,f], and David Zilberman[e]

[a]Charles University, Prague

[b]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic

[c]University of Economics, Prague

[d]CERGE-EI

[e]University of California in Berkeley

[f]Centre for Applied Macroeconomic Analysis (CAMA), ANU

November 17, 2012

Abstract

We analyze the relationships between biodiesel, ethanol and related fuels and agricultural commodities with a use of minimal spanning trees and hierarchical trees and Granger causality. We construct this trees for different frequencies (weekly, monthly and quarterly). We find that in short-term, both ethanol and biodiesel are very weakly connected with the other commodities. In medium and long term, the network structure becomes more interesting. We especially concentrate on the links and comovements between the commodities for different stages of the market based on the level of prices of important commodities or groups of commodities. Such approach in general confirms the separation of network into fuels and food branches for majority of analyzed cases.

Keywords: biofuels, networks, causality

JEL Codes: C22, C38, Q42

*Ladislav Kristoufek, kristoufek@ies-prague.org, Karel Janda, Karel-Janda@seznam.cz, David Zilberman, zilber11@berkeley.edu.

1 Introduction

After the oil crisis of 1970s, biofuels became of high interest as a possible replacement for fossil liquid fuels used in transportation. Increased interest in climate and environmental issues in last three decades also contributed to the popularity of biofuels as alternative fuels. Global production of biofuels experienced a rapid increase since then, especially during the last decade.

Biofuels represent a wide range of fuels which are in some way derived from biomass. Therefore the wide definition of biofuels covers solid biomass, liquid fuels and various biogasses. In this paper, we focus on the two most popular biofuels – ethanol and biodiesel.

Ethanol is mainly produced from crops rich in sugar and starch like sugarcane and corn. Other major crops being used include wheat, sorghum, sugar beet, and cassava. In contrast to ethanol, biodiesel is produced from oilseed crops like soybean, rapeseed, and oil palm. The main aim of our research is to analyze connections between these two biofuels and related commodities (both energy-related and food-related) with a use of minimum spanning trees (MST) and hierarchical trees (HT) to uncover the most important connections in the network of commodities. MST and HT are methodologically simple approaches using only simple correlations as a starting point with no prior assumptions. The methods have been applied several times on various types of assets and systems – stocks and stock indices [2, 27], foreign exchange rates [11], import/export networks [12], interest rates [25], and commodities [26].

We focus on analysis of biofuels networks and their structure during various phases of the market. As a follow-up to our previous research, where we analyzed the network before and after the food crisis of 2007/2008 [15], we focus on situations when various prices of various parts of the system are high or low to see possible non-linearity of the links between relevant commodities. Moreover, we also analyze these connections at different frequencies (week and month) to observe possible short-term and medium-term effects.

The rest of the paper is structured as follows. In Section 2, we present a brief literature review of empirically oriented papers dealing with biofuels. In Section 3, we describe the basic notions of the used methodology. In Section 4, the data choice and description is given. Section 5 presents the results of our analysis. Section 6 concludes.

2 Brief literature review

In this section, we shortly review the most recent research on links between biofuels and other commodities.

Tyner [28] argues that since 2006, the ethanol market has established a link between crude oil and corn prices which was nonexistent previously. Focusing on the cointegration between oil, ethanol and feedstocks, Serra and co-authors [24, 23] analyze the US and Brazilian markets. They find the existence of a long term equilibrium relationship between these commodities with ethanol deviating from this equilibrium in the short term for the USA. For Brazil, the authors find that sugar and oil prices are exogenously determined and focus their attention on the response of ethanol prices to changes in these two exogenous drivers finding that ethanol prices respond relatively quickly to sugar price changes and more slowly to oil prices.

While ethanol is the most important biofuel on US and Brazilian markets, in Europe biodiesel plays the prominent role. It is not only in Germany, but also in other European countries. For example Hassouneh *et al.* [9] find on Spanish data that biodiesel, sunflower and crude oil prices are in a long-run, equilibrium relationship. They discover that biodiesel is the only variable that adjusts to deviations from this long-run relationship. Local linear regression techniques then show that the speed of adjustment of biodiesel prices is faster when biodiesel is relatively cheap than when it is expensive. This regime difference between behavior under low and high prices is similar to our approach when we distinguish among the different topologies of biofuels network depending on price levels.

Possibility to add biofuels to the portfolio and related portfolio selection issues have been analyzed and discussed as well [29, 1]. Zhang *et al.* [30] further examine volatility in ethanol and commodity prices using cointegration, vector error corrections (VECM) and multivariate generalized autoregressive conditional heteroskedascity (mGARCH) models to find that recently, there are no long-run relations among fuels (ethanol, oil and gasoline) and agricultural commodities (corn and soybean). The same authors [31] then focus on Granger-causality and find no direct long-run price relations between fuel and agricultural commodity prices, and only limited if any direct short-run relationships.

Du and co-authors [7] investigate the spillover of crude oil volatility to agricultural markets applying stochastic volatility models and Bayesian Markov Chain Monte Carlo methods. They find that the spillover effects are insignificant between November 1998 and October 2006, but for October 2006-January 2009 period, they uncover significant volatility spillover from the crude oil market to the corn market. Kristoufek *et al.* [15] examine the biofuels network with a use of minimal spanning trees and hierarchical trees for the periods before and during/after the food crisis of 2007/2008, when we experienced unprecedentedly high prices of almost all agricultural commodities, to show that links indeed differ in the two periods.

In general, there are many studies analyzing biofuels network with many different approaches (for more detailed reviews, see [10, 22, 32]). However, as is also evident from the short review we give, the results are quite ambiguous. To overcome this, we employ a straightforward methodology of MST and HT which are model-free while analyzing very broad system of biofuels and related commodities (in comparison to other biofuels related studies). By doing so, we should be able to find the most basic connections between the commodities. We also complement our basic network topology approach with additional analysis of causal relationships among the commodities whose close relationship was uncovered through MST/HT approach. In this we follow the Granger causality approach to biofuels related price transmission presented by [4, 3, 19, 20, 21] and others.

3 Methodology

3.1 Distance measure

Correlations between pairs of assets are the very basic measures of interdependence in the whole system. For a pair of assets i and j with values X_i and X_j and $i, j = 1, \ldots, T$, the sample correlation coefficient $\widehat{\rho_{ij}}$ is defined as

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

where $\bar{X}_i = \frac{\sum_{t=1}^{T} X_{it}}{T}$ and $\bar{X}_j = \frac{\sum_{t=1}^{T} X_{jt}}{T}$ are respective averages. Correlation coefficient ρ_{ij} ranges between -1 and 1 for perfectly negatively and perfectly positively correlated series, respectively. For $\rho_{ij} = 0$, we have a linearly uncorrelated pair. Asymptotic stationarity of the series is assumed.

For large systems, the number of pairs can become unbearable while some or sometimes majority of the links between pairs are not important for the characteristics of dependencies in the network. Note that for a portfolio of N assets, we obtain $N(N-1)/2$ pairs of correlations. Mantegna [17] proposed to transform the correlation coefficients into distance measures and used these to describe hierarchical organization of the group of analyzed assets. Proposed distance measure

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

fulfills the three axioms of a metric distance:

- $d_{ij} = 0$ if and only if $i = j$;
- $d_{ij} = d_{ji}$;
- $d_{ij} \leq d_{ik} + d_{kj}$ for all k

The distance d_{ij} ranges between 0 and 2 – $d_{ij} \to 0$ means that the pair is strongly correlated, $d_{ij} \to 2$ implies strongly anti-correlated pair and $d_{ij} = \sqrt{2}$ characterizes an uncorrelated pair.

3.2 Minimal spanning tree and hierarchical tree

From the distance measures, we can extract the most important connections in the network and construct the minimal spanning tree (MST). MST consists of the $N-1$ most important connections in the system while remaining connected. In the process, $N(N-1)/2$ correlations are reduced to MST with $N-1$ links. The procedure is in detail described in [17]. In short, the correlation matrix $\mathbb{C}$ is transformed into a distance matrix $\mathbb{D}$, erasing the diagonal elements (zero distances). The closest pair of assets creates the first two nodes in the network connected by a link (with a weight equal to the distance d_{ij}). The second closest pair then creates the second pair of nodes. At this point, if a node from the second pair is already a part of MST, the new node is connected to the existing pair. This procedure is repeated until $N-1$ links are found without closing the network or creating closed loops. If the link would create a loop, it is discarded. Kruskal's algorithm [16] is used for the construction of MST.

To obtain more information about the links and mainly clusters in the system, hierarchical trees (HT) are constructed from the MSTs. Hierarchical trees have been used to show that stocks form clusters based on the industrial branches [17, 26] and that foreign exchange rates create clusters with respect to the geographical situation [18, 13, 11]. To construct HT, we need to determine the subdominant ultrametric distance matrix $\mathbb{D}^*$, which consists of subdominant ultrametric distances d^*_{ij}. The distance is found as a maximal weight of the link on a path which needs to be taken to move from node i to node j in the minimal spanning tree. This way, we can identify clusters of assets which have the same d^*_{ij} to some other asset or assets. Detailed description of the procedure is given by Mantegna [17].

3.3 Statistical significance

Minimal spanning and hierarchical trees can be unstable and for a specific system, links can emerge randomly or as an anomaly. To control for this, we apply a bootstrapping technique proposed by Tumminello *et al.* [27] specifically for MST and HT analysis. The application is very straightforward. MST and HT are firstly constructed for the original series. A new multivariate series is then resampled with repetition from the original series keeping the length of the series. Therefore, the core of correlations remains undistorted. New MST and HT are constructed for the bootstrapped series and the links are recorded. This is repeated 1,000 times so that we can distinguish between stable and unstable links [13].

4 Data

We analyze the weekly data of Brent crude oil (CO), ethanol (E), corn (C), wheat (W), sugar cane (SC), soybeans (S), sugar beets (SB), consumer biodiesel (BD), German diesel and gasoline (GD and GG), and the U.S. diesel and gasoline (UD and UG) from 24.11.2003 to 28.2.2011. Commodities data have been obtained from Bloomberg database and except for the biofuels, we use the 1-month futures prices. For biodiesel and ethanol, we use spot prices. Gasoline and diesel prices were obtained from the U.S. Energy Information Administration. We keep both the fuels of the USA and the EU (Germany as the biggest economy of the EU) to find potential geographical patterns and also to control for the fact that ethanol is mainly the US (and also Brazilian) domain and biodiesel is rather the EU domain. Due to illiquidity of the analyzed biofuels, we analyze weekly data. Descriptive statistics for the log-returns (Monday-Monday closing prices) are presented in Table 1. We observe that all the retail fuels are positively skewed and all the others assets but corn are negatively skewed. Tests for stationarity are summarized in Table 2 and show that all the analyzed log-returns series are asymptotically stationary so that we can proceed with our analysis.

Apart from analyzing the network between commodities for the whole sample period, we also examine connections between series for different phases of the market. To do so, we split the series according to a simple rule – we choose an important commodity or a group of commodities and split the series into two parts, one comprising of returns in a case the price was below the median price and the other with the returns when the price was above the median price. We use the median to ensure that the series are of the same length (a simple average is too sensitive to the extreme values and might result in two series of very different lengths, which would in turn be hardly comparable). From the set of variables and their combinations, we choose four – crude oil, feedstock, fuels and sugars. Therefore, we arrive at eight different stages of the market – high crude oil prices, low crude oil prices, high feedstock prices (when both wheat and corn are above their medians), low feedstock prices (when both wheat and corn are below their medians), high fuels prices (when the average of German gasoline, German diesel, US gasoline and US diesel is above its median), low fuels prices (when the average of German gasoline, German diesel, US gasoline and US diesel is below its median), high sugar prices (when both sugarcane and sugar beets are above their medians) and low sugar prices (when both sugarcane and sugar beets are below their medians). For each of the phases, we analyze the weekly and monthly series (quarterly frequency had to be omitted due to too few observations). Since the separate groups are no longer treated as the time series, we do not need to bother with the stationarity issue.

5 Results and discussion

5.1 Whole sample

Let us first focus on the MST for the whole period (Fig. 1). It is clearly visible that the minimal spanning tree is formed from two parts – a food part (SC, SB, W, C, S) and a fuels part (CO, GD, GG, UG, UD, E, BD). In the MST charts, we also show the distances between nodes (regular font) as well as a bootstrapped value (italics in brackets). The bootstrapped value represents the proportion of times when the specific link has been present in the bootstrapped MST. For example, the value of 0.783 for S—CO link means that out of 1,000 bootstrapped realization, the S—CO link has been found in 783 final MSTs. Using these values, we can comment on a strength or a stability of a link in the MST. In the food part of the MST, we observe a triple W—C—S and a pair SC—SB which have been found in all bootstrapped realizations. These links are thus very stable. The connection between the triple and the pair is quite weaker. We can see similarly strong connections in the fuels part of the MST, mainly for quadruple GD—GG—UG—UD which has been found in almost all the bootstrapped cases. Both biofuels are linked to the US fuels. Relatively low bootstrapped value for CO—GD link is caused mainly by the fact that crude oil is correlated to GG, GD, UD and UG at similar levels so that the links alter between the four in the bootstrapped cases.

Very similar results can be read from the HT[1]. Here, we can see that there are several clusters – a fuels cluster, a sugar cluster and a fodder cluster. The other commodities – crude oil, ethanol and biodiesel – are quite far from these clusters and thus do not interact much in the short term. Importantly, the biofuels are quite remote from the rest of the network, which can be interpreted in a way that in a short term horizon, the behavior of these biofuels is not dependent on the other analyzed commodities.

When we look at the relationships between commodities at lower frequencies, both MSTs and HTs are getting more interesting. The core of the connections remains the same – we still have the three clusters. However, the behavior of the biofuels changes. Ethanol becomes more connected with the food part and biodiesel with the fuels part. Interestingly, the whole network practically splits into two branches for the lowest frequency (a quarter of a year) – one branch contains all the standard fuels, crude oil and biodiesel and the other branch includes all the analyzed food and ethanol.

To summarize the most important findings for ethanol and biodiesel returns with respect to different frequencies, we can say that in short-term, both of these are very weakly connected with the other commodities. Moreover, there is no clear inclination to either of

[1] Hierarchical trees were constructed using *stats* package in R 2.15.1.

fuels or food parts of the network. In medium and long-term, biodiesel becomes connected to the fuels section of the system, whereas ethanol gets more connected to the food branch of the system. It seems that the longer the horizon, the more ethanol is connected with the sugars. Therefore, it takes quite long before these two affect one another. In economic terms, the effects on price of SC and SB take a long time before they influence the price of ethanol. Unfortunately, the MST and HT analysis is not capable to find the direction of this effect, i.e. whether the effect comes from food to ethanol or the other way around. However, we may speculate that if the effect was short-term, it is more likely that the direction is from ethanol to food because this effect can be very quick. On the other hand, the production cycle of the biofuels take much longer than speculations on commodities markets.

5.2 Market phases

5.2.1 Crude Oil phases

MST and HT for the phases based on the crude oil prices are shown in Fig. 2. We first focus on the situations when the crude oil prices are above the median prices (Fig. 2a-d). In a short-term (Fig. 2a), the most stable connections are again between GG—GD, UG—UD, SB—SC and S—C—W. These three pairs and one triple are not only highly correlated but also very stable (based on the bootstrap results). Even though the connection CO—GG is quite close (correlated), it is rather unstable. The most outlying commodity is biodiesel with a distance from UD of 1.313 (keep in mind that an uncorrelated series have a distance of $\sqrt{2} \approx 1.414$). In medium term (Fig. 2c), there are two interesting differences – soybeans are separated from the C—W pair and ethanol becomes very separated from the fuels part of the tree with a distance of 1.4664, i.e. negatively correlated. This high distance strongly separates the food part and the fuels part of the system indicating that these two branches are negatively correlated, even though very weakly.

More interesting implications can be found when looking at the hierarchical trees. In short term (Fig. 2b), there are no interesting results and both biofuels remain far from the other commodities while the rest of the network is separated into three parts – fuels, feedstock and sugars. In medium term (Fig. 2d), the results become more interesting as the system splits into two (as was obvious from the corresponding MST) branches – fuels and food. Biodiesel is a part of the fuels branch and ethanol of the food part. This implies that in a market with high crude oil prices, the prices of biodiesel are mainly driven by other fuels and the prices of ethanol are mainly driven by the changes in the food commodities (mainly in soybeans and sugars).

In the opposite phase of the market (with low crude oil prices, Fig. 2e-h), the results are again more interesting for a medium term. Nevertheless, for the weekly frequency, we again observe that both biofuels are very far from the rest of the system. Interestingly, crude oil is far from the other fuels. For a medium term, biodiesel becomes very connected to the fuels branch of the system (with a distance from crude oil equal to 0.9391, which is also very stable with a bootstrap value of 0.962). On the other hand, ethanol remains relatively far from the other parts of the network. Moreover, crude oil becomes very close to the rest of the fuels section.

From the biofuels perspective, prices of ethanol become driven by the commodities from the food branch in medium term when the prices of crude oil are high. Otherwise, ethanol remains rather lowly correlated with the system. On the other hand, biodiesel becomes strongly influenced by the crude oil movements when the prices of crude oil are low. Crude oil also shows interesting dynamics with respect to the retail fuels, which implies that the prices of retail fuels adjust faster when the crude oil prices are high than when these are low.

5.2.2 Feedstock phases

The effect of high and low corn and wheat prices on the network between commodities is shown in Fig. 3. We again start with the effect of high prices. In Fig. 3a-d, MST and HT for weekly and monthly data are illustrated. In a short term, we again find strong and stable connections GG—GD, UG—UD, SC—SB, S—C—W. However, we also see that ethanol is close to the triplet S—C—W with quite stable connection (with a bootstrap value of 0.872). Also, crude oil is closer to the food commodities than to the fuel commodities. Biodiesel is again very far from the rest of the network with an unstable connection (a bootstrap value of 0.226). In medium term, MST is quite similar – we observe strong connections between retail fuels, feedstock and sugars. However, the whole system is strongly separated into two sections with a very weak and unstable connection CO—W with a distance of 1.3119.

The connections and clustering between commodities are more obvious from the hierarchical trees. In short term, we observe a cluster of retail fuels, a cluster of food products (containing crude oil and ethanol) and a separated branch of biodiesel. This implies that the fodder commodities are more sensitive to the changes in crude oil prices when the prices of the feedstock are high. Moreover, ethanol has quite strong and stable connection to the feedstock commodities and is thus influenced and correlated with the changes in feedstock commodities and crude oil. In medium term, the structure of the network and the hierarchical tree resembles the structure for the whole sample (Fig. 1d) – two well separated branches of fuels and food. Biodiesel is connected to the fuels branch (even though relatively weakly) and ethanol is well connected to the food branch (specifically strongly connected to corn). Most importantly, ethanol is strongly connected (with stable connections) to feedstock commodities in both short and medium term when the prices of feedstock are high.

When the prices of feedstock are low (Fig. 3e-h), the structure of the network changes greatly for the medium term. For the short term, we observe the same three couples and one triplet as before and very weak connections of ethanol and biodiesel to the rest of the system. Moreover, the connections between the couples and the triplet are rather unstable. In medium term, the structure of the HT becomes quite messy – SB—SC is connected with GG—GD which is in turn connected to C—W—S before it is connected to UD—UG—CO—BD. However, the connection between the sugars and German retail fuels is rather unstable (a bootstrap value of 0.327). More so for the connection between the triplet S—C—W and GD—GG with an unstable bootstrap value of 0.300. Therefore, the most important outcome for the medium term with low feedstock prices remains the outlying ethanol (with a distance of 1.2513 from the rest of the network) and the fact that biodiesel is well and quite stably connected to crude oil and retail fuels.

In summary, in the phase of the market when corn and wheat prices are high, ethanol is well connected and correlated with feedstock prices in both short and medium term while in the phase of low fodder prices, it is very weakly connected to the rest of the system. For biodiesel, we observe the same behavior for both phases which is also very similar to the behavior for the same sample, i.e. very low correlation in short term and a fair connection to the fuels branch of the system in medium term.

5.2.3 Fuels phases

We now focus on the phases of the market with respect to the average price of all retail fuels. Minimal spanning trees and hierarchical trees are illustrated in Fig. 4. In the situation when the retail fuel prices are high and the time horizon is short (Fig. 4a-b), we observe standard strong and stable connections UD—UG, GD—GG, S—C—W and SB—SC. Moreover, crude oil is close to the fodder triplet, while this connection is very stable with a bootstrap value of 0.985. On the other hand, biodiesel and ethanol are rather far from the other commodities. Standard picture is supported by HT in Fig. 4b , where we observe standard three branches of fuels, sugars and fodder. Ethanol and biodiesel are far from the rest of the system. Moreover, crude oil is a part of the food branch and rather far from the fuels branch (distance of 1.1134 with a bootstrap value of 0.593, i.e. modestly stable). This implies that in short term, crude oil and fuels (at high prices) are only weakly correlated. This finding is different compared to the results of high crude oil prices, where we observed strong connection between crude oil and retail fuels when the crude oil prices were high.

In medium term (Fig. 4c-d), we observe rather different results. Importantly, the connections between commodities are less stable. Also, the standard feedstock triplet is separated and we find a strong connection only between corn and wheat. The other standardly strong connection between retail fuels and sugars remain. Interestingly, ethanol becomes well connected to the GG—GD pair. On the other hand, biodiesel is very far from the rest of the network and is even negatively correlated with the rest of the system. This is a very interesting result because in other cases, we found ethanol to be more connected to the food branch of the system.

When we look at the opposite market phase (low fuel prices), we find standard stable connections between retail fuels, sugars and feedstock for both short and medium term (Fig. 4e-h). Biodiesel is a part of the fuels branch of the system while the correlation is very similar for both short and medium term. On the other hand, ethanol is well connected with the fuels branch in the short term but becomes very far from the rest of the system in the medium term. Crude oil is connected to the fuels part of the network for both short and long term, while the distance is much higher for the short term.

Summarizing the behavior of the commodities in different phases of the market with respect to retail fuels, we find several interesting facts. When the retail fuel prices are high, biodiesel is practically outside of the system compared to ethanol which becomes sensitive to the movements in fuels and crude oil in the medium term. And when the fuel prices are low, biodiesel is correlated with the fuels in the system in both short and medium term compared to ethanol which is close to the fuels branch only in short term. Interestingly, crude oil is not well connected with other fuels in short term when the fuels prices are high. If we add this up together with the results of crude oil oil market phases, we find that fuels prices are well correlated with the high crude oil prices but not the other way around. This implies that when the crude oil prices are high, retail fuels are also expensive. On the other hand, when fuels prices are high, the fuels are not necessarily well connected to crude oil. The effect of expensive or cheap crude oil prices on retail fuels is then asymmetric.

5.2.4 Sugars phases

Last investigated stage of the market is separating the system with low or high sugars prices (Fig. 5). When the prices of sugars are high and in a short term, we again observe very stable and strong connections between retail fuels, sugars and fodder. Biodiesel is far from the system (with the distance of 1.3381) and ethanol is closest to soybeans (with distance of 1.1675). From the clusters perspective, we observe something very non-standard. While ethanol and biodiesel are quite standardly well apart from the rest of the system, the retail fuels are very geographically distinct. Even though the connection between German and US fuels is very stable (a bootstrap value of 0.897), the distance is rather high (1.0824) compared to the other scenarios discussed previously.

The situation changes for the medium term, where we observe rather usual separation into three branches of the system, i.e. fuels, feedstock and sugar branches. While biodiesel remains far from the rest of the system (with a distance of 1.4244, i.e. practically uncorrelated), ethanol is well connected to the fuels branch (linked to German diesel with a distance of 1.0264).

Looking at the other situation on the market (low sugars prices), we find some interesting results. The retail fuels are again strongly and stably connected to each other, as well as sugars and the fodder triplet remains well and stably linked as well. Interestingly, both ethanol and biodiesel are quite well connected to the fuels branch of the system in the short term. Surprisingly, crude oil is connected to the fodder branch of the system. However, this connection is rather unstable with a bootstrap value of 0.266. In the medium term, the situation changes in some aspects. Most importantly, crude oil is now well connected to the fuels branch as expected. Biodiesel remains linked with the fuels part of the network but ethanol switches to the other part of the system and is connected to sugar commodities.

To summarize the results for different phases of the market based on sugars prices, we find that biodiesel remains unattached to the system when the sugars prices are high in both short and medium term. On the other hand, ethanol is influenced by the fuels when the prices of sugars are high in the medium term. When the sugars prices are low, ethanol is correlated with their changes in medium term.

5.3 Causality

Even though our analysis of the minimum spanning and hierarchical trees uncovered interesting (and statistically significant) connections, these don't give any information about causality between biofuels and related commodities, which is, however, of interest from economic and social perspective. The main reason is the potential squeeze-out effect between biofuels and related agricultural commodities. Increasing prices (and higher revenues) of biofuels might motivate entrepreneurs and firms to move to the biofuels industry and, if these are already in the agriculture industry, they may replace previously cultivated crops with biofuels-producing commodities and thus increasing the prices of the former. Also, the already grown biofuels-producing agricultural crops might be further used for the biofuels production rather than for food production cycle and thus increasing the food prices, which brings noticeable social burden. On the other side, increasing prices of agricultural commodities can increase prices of the related biofuels as these are their producing factors, which comes from the very basic economic theory. Causal relationships among agricultural commodities and fuels are also relevant for the analysis of environmental effects of biofuel production and the evaluation of biofuels related policies from the climate change mitigation point of view [5, 6, 14].

To check which of these two possibilities outlined in the previous paragraph is present in the analyzed dataset, we apply Granger causality methodology [8]. Note that the methodology can be only used on the original series as the "market phases" series are not continuous and thus cannot be treated as the time series. Based on the results already presented for the whole sample, we analyze the causality only for the four week and the twelve week frequency as there is no strong relationship between any of the biofuels and related commodities at the weekly frequency. From Fig. 1, it is evident that ethanol forms a cluster with soybeans, corn and wheat (for the monthly frequency), and sugarcane and sugar beets (for the quarterly frequency), and biodiesel forms a cluster with the fuels (for both frequencies). The results for both frequencies Granger tests and the summed estimated coefficients for relevant pairs are summarized in Table 3.

For ethanol, we find that the increasing prices of both corn and wheat strongly and positively influence the prices of ethanol in both medium and long term, while the effect of corn is stronger. Other pairs show no significant Granger-causal relationship, which includes the finding that ethanol does not Granger-cause any of the included agricultural commodities. For biodiesel, we find that its price is strongly and again positively influenced by increasing prices of German diesel and crude oil in both medium and long horizon. In the other direction, there is no significant relationship. Note that we analyzed the relationship between biodiesel and only German diesel and crude oil since the other fuel commodities are connected to these two rather than directly to biodiesel, which is mainly produced in the EU (thus German diesel rather than the US diesel). Also, biodiesel is obviously connected to diesel rather than to gasoline.

Our Granger causality results presented in the previous paragraph therefore show that the prices of the analyzed biofuels are boosted by the increasing prices of their producing factors and not the other way around.

6 Conclusions

We analyzed the relationships between biodiesel, ethanol and related fuels and agricultural commodities with a use of minimal spanning trees and hierarchical trees. To distinguish between short term and long term effects, we constructed the trees for different frequencies (weekly, monthly and quarterly). We found that in short-term, both biofuels are very weakly connected with the other commodities. In medium and long term, the network structure becomes more interesting. The system splits into two well separated branches – a fuels part and a food part. Biodiesel tends to the fuels branch and ethanol to the food branch.

When we analyzed the connections between commodities for different stages of the market, we uncovered several interesting results. First, prices of ethanol become driven by the commodities from the food branch in medium term when the prices of crude oil are high. Second, biodiesel becomes strongly influenced by the crude oil movements when the prices of crude oil are low. Third, when corn and wheat prices are high, ethanol is well connected and correlated with feedstock prices in both short and medium term. Fourth, biodiesel is correlated with the fuels in the system in both short and medium term compared to ethanol which is close to the fuels branch only in short term when the retail fuels prices are low. Fifth, ethanol is influenced by the fuels when the prices of sugars are high in the medium term, and when the sugars prices are low, ethanol is correlated with their changes in medium term. And sixth, the prices of the analyzed biofuels react positively to the increasing prices of their producing factors and not the other way around.

Acknowledgements

The authors acknowledge the support from Energy Biosciences Institute at University of California, Berkeley (EBI). Karel Janda acknowledges research support provided during his long-term visits at University of California, Berkeley and Australian National University. Our research was supported by the grants P402/11/0948 and 402/09/0965 of the Czech Science Foundation, grant 118310 and project SVV 265 504 of the Grant Agency of the Charles University and by institutional support grant VSE IP100040. The views expressed here are those of the authors and not necessarily those of our institutions. All remaining errors are solely our responsibility.

References

[1] Rob Bailis, Barbara Sophia Koebl, and Mark Sanders. Reducing fuel volatility – an additional benefit from blending bio-fuels? Discussion Paper 11-01, Tjalling C. Koopmans Research Institute, February 2011.

[2] G. Bonanno, G. Caldarelli, F. Lillo, S. Micciche, N. Vandewalle, and R. Mantegna. Networks of equities in financial markets. *European Physical Journal B*, 38:363–371, 2004.

[3] Pavel Ciaian and D'Artis Kancs. Food, energy and environment: Is bioenergy the missing link? *Food Policy*, 36(5):571–580, October 2011.

[4] Pavel Ciaian and D'Artis Kancs. Interdependencies in the energy-bionergy-food price systems: A cointegration analysis. *Resource and Energy Economics*, 33(1):326–348, January 2011.

[5] Harry de Gorter and David R. Just. The economics of a blend mandate for biofuels. *American Journal of Agricultural Economics*, 91(3):738–750, August 2009.

[6] Dusan Drabik, Harry de Gorter, and David R. Just. The implications of alternative biofuel policies on carbon leakage. Working Paper 2010-22, Cornell University, December 2010.

[7] Xiaodong Du, Cindy L. Yu, and Dermot J. Hayes. Speculation and volatility spillover in the crude oil and agricultural commodity markets: A Bayesian analysis. *Energy Economics*, 33(3):497–503, 2011.

[8] C. Granger. Investigating casual relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

[9] Islam Hassouneh, Teresa Serra, Barry K. Goodwin, and Jose M. Gil. Non-parametric and parametric modeling of biodiesel, sunflower oil, and crude oil price relationships. *Energy Economics*, 34(5):1507–1513, September 2012.

[10] Karel Janda, Ladislav Kristoufek, and David Zilberman. Biofuels: Policies and impacts. *Agricultural Economics*, 58(8):367–371, August 2012.

[11] W. Jang, J. Lee, and W. Chang. Currency crises and the evolution of foreign exchange market: Evidence from minimum spanning tree. *Physica A*, 390:707–718, 2011.

[12] E. Kantar, B. Deviren, and M. Keskin. Hierarchical structure of Turkey's foreign trade. *Physica A*, 390(20):3454–3476, 2011.

[13] M. Keskin, B. Deviren, and Y. Kocakaplan. Topology of the correlation networks among major currencies using hierarchical structure methods. *Physica A*, 390:719–730, 2011.

[14] Madhu Khanna, Christine L. Crago, and Mairi Black. Can biofuels be a solution to climate change? the implication of land use change related emissions for policy. *Interface Focus: The Royal Society Journal*, 1(2):233–247, April 2011.

[15] Ladislav Kristoufek, Karel Janda, and David Zilberman. Correlations between biofuels and related commodities before and during the food crisis: A taxonomy perspective. *Energy Economics*, 34(5):1380–1391, September 2012.

[16] J. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50, 1956.

[17] R.N. Mantegna. Hierarchical structure in financial markets. *European Physical Journal B*, 11:193–197, 1999.

[18] T. Mizuno, H. Takayasu, and M. Takayasu. Correlation networks among currencies. *Physica A*, 364:336–342, 2006.

[19] Jan Pokrivcak and Miroslava Rajcaniova. Crude oil price variability and its impact on ethanol prices. *Agricultural Economics – Czech*, 57(8):394–403, August 2011.

[20] Miroslava Rajcaniova, Dusan Drabik, and Pavel Ciaian. International interlinkages of biofuel prices: The role of biofuel policies. Presentation at AAEA Meeting, Pittsburgh, July 2011.

[21] Miroslava Rajcaniova and Jan Pokrivcak. The impact of biofuel policies on food prices in the European Union. *Journal of Economics (Ekonomicky casopis)*, 59(5):459–471, 2011.

[22] Teresa Serra and David Zilberman. Biofuel-related price transmission literature: A review. Unpublished Manuscript, 2012.

[23] Teresa Serra, David Zilberman, and Jose M. Gil. Price volatility in ethanol markets. *European Review of Agricultural Economics*, 38(2):259–280, 2011.

[24] Teresa Serra, David Zilberman, Jose M.Gil, and Barry K. Goodwin. Nonlinearities in the U.S. corn-ethanol-oil-gasoline price system. *Agricultural Economics*, 42(1):35–45, January 2011.

[25] B. Tabak, T. Serra, and D. Cajueiro. The expectation hypothesis of interest rates and network theory: The case of Brazil. *Physica A*, 388:1137–1149, 2009.

[26] B. Tabak, T. Serra, and D. Cajueiro. Topological properties of commodities networks. *European Physical Journal B*, 74:243–249, 2010.

[27] M. Tumminello, T. Di Matteo, T. Aste, and R. Mantegna. Correlation based networks of equity returns sampled at different time horizons. *European Physical Journal B*, 55:209–217, 2007.

[28] Wallace E. Tyner. The integration of energy and agricultural markets. *Agricultural Economics*, 41(s1):193–201, November 2010.

[29] Zibin Zhang, Luanne Lohr, Cesar Escalante, and Michael Wetzstein. Mitigating volatile U.S. gasoline prices and internalizing external costs: A win-win fuel portfolio. *American Journal of Agricultural Economics*, 90(5):1218–1225, December 2008.

[30] Zibin Zhang, Luanne Lohr, Cesar Escalante, and Michael Wetzstein. Ethanol, corn, and soybean price relations in a volatile vehicle-fuels market. *Energies*, 2(2):320–339, 2009.

[31] Zibin Zhang, Luanne Lohr, Cesar Escalante, and Michael Wetzstein. Food versus fuel: What do prices tell us. *Energy Policy*, 38(1):445–451, January 2010.

[32] D. Zilberman, G. Hochman, D. Rajagopal, S. Sexton, and G. Timilsina. The impact of biofuels on commodity food prices: Assessment of findings. *American Journal of Agricultural Economics*, doi:10.1093/ajae/aas037:Forthcoming, 2012.

Table 1: Descriptive statistics

	mean	min	max	SD	skewness	ex. kurtosis
Crude Oil	0.0035	-0.2971	0.2010	0.0500	-0.7501	4.2687
Ethanol	0.0011	-0.2097	0.2085	0.0515	-0.1446	2.8889
Corn	0.0029	-0.1790	0.2028	0.0482	-0.2125	1.5812
Wheat	0.0019	-0.1699	0.1476	0.0488	0.0230	0.4472
Sugar cane	0.0043	-0.1921	0.2008	0.0514	-0.1910	1.1557
Soybeans	0.0016	-0.2989	0.2333	0.0482	-1.2682	7.4407
Sugar beets	0.0035	-0.1546	0.1151	0.0415	-0.3639	0.9818
Biodiesel	0.0014	-0.0618	0.0760	0.0156	0.2292	3.9239
German diesel	0.0028	-0.1206	0.1420	0.0437	0.2026	0.3521
US diesel	0.0026	-0.1099	0.1379	0.0260	0.3980	4.4957
German gasoline	0.0026	-0.1806	0.2256	0.0508	0.3609	1.9525
US gasoline	0.0024	-0.1049	0.1831	0.0276	0.1612	6.1811

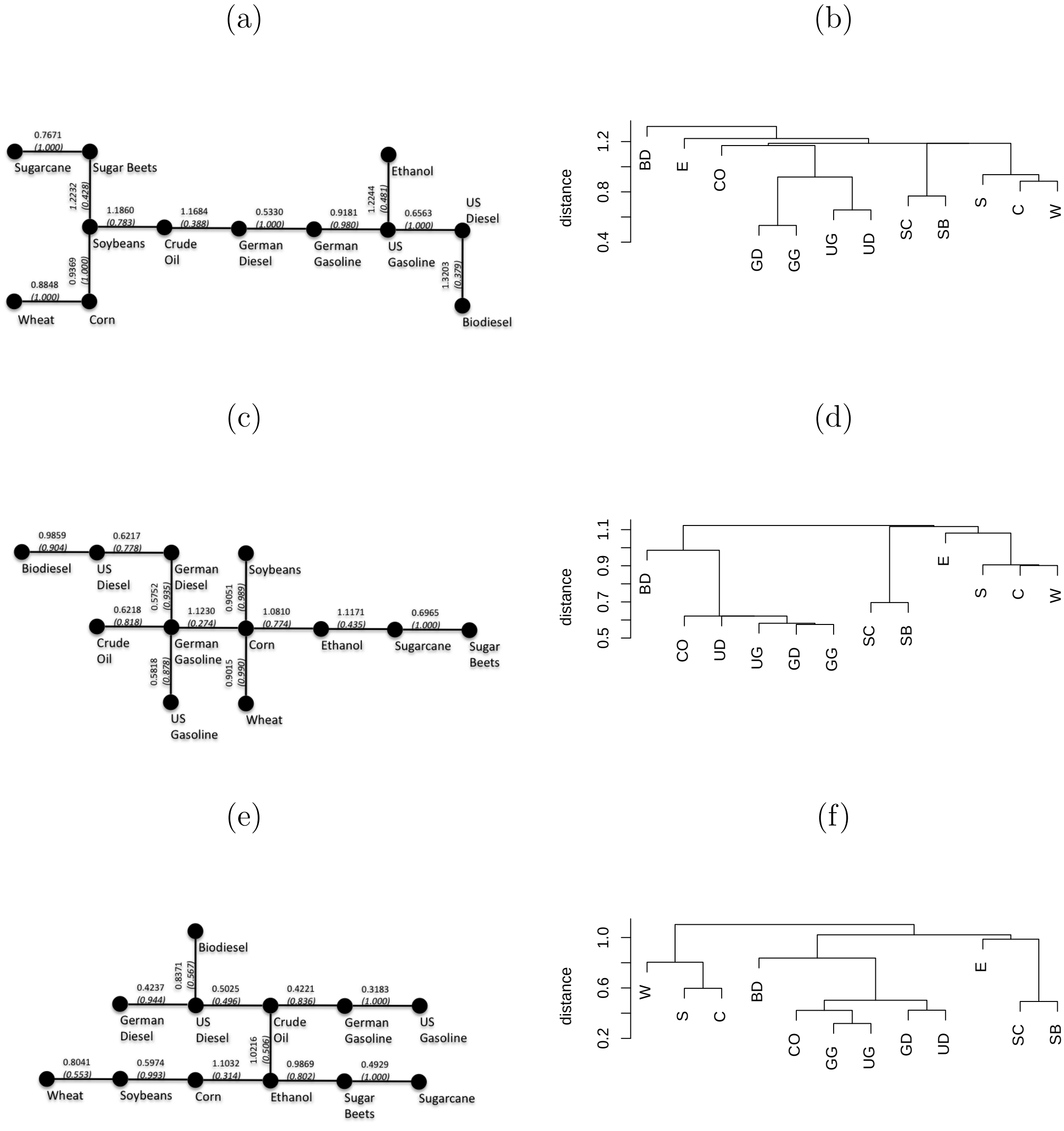

Figure 1: *Minimal spanning trees (first column) and hierarchical trees (second column) for network of returns and different frequencies (from the top – one week, one month and one quarter)*

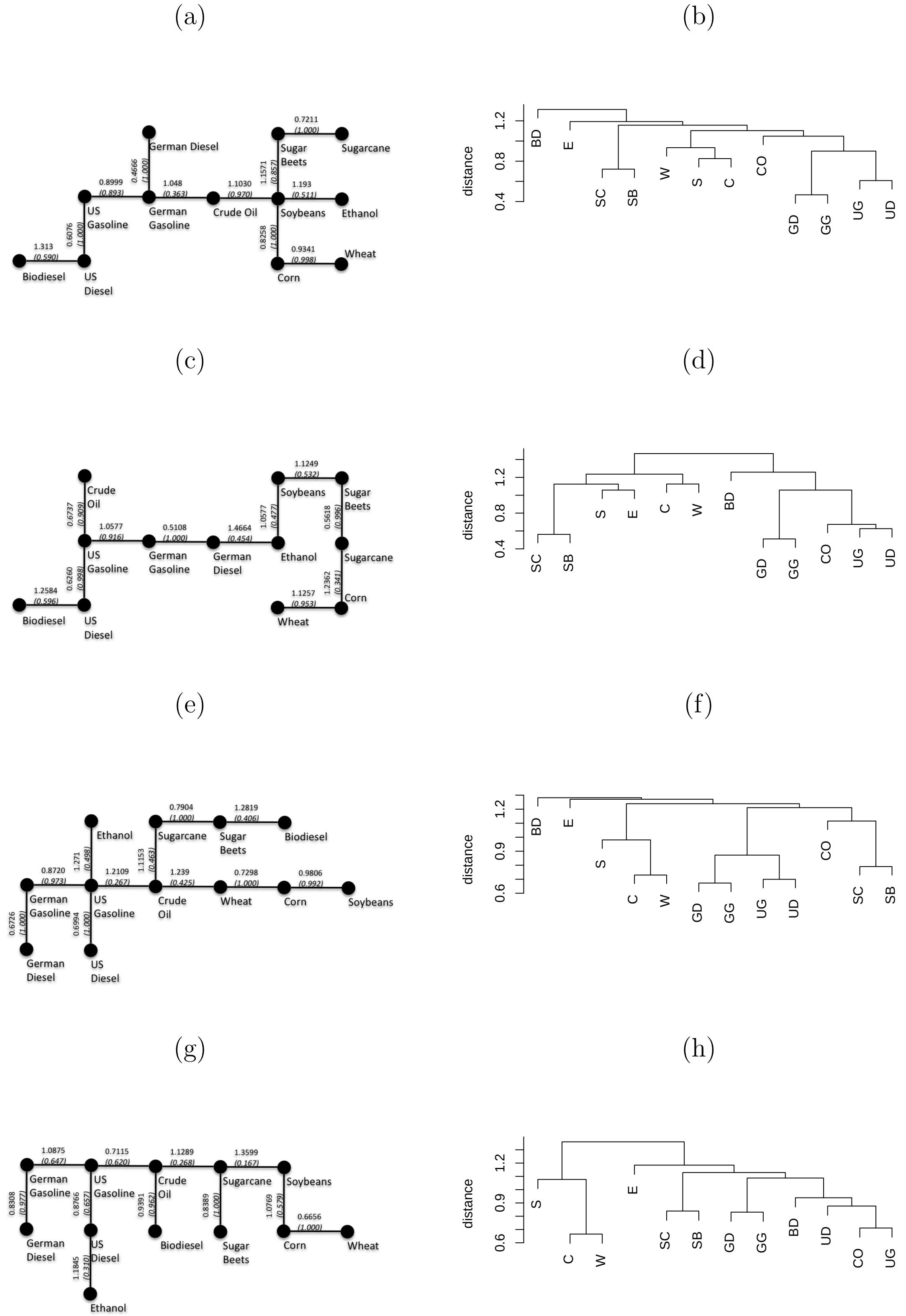

Figure 2: *Crude Oil.* Results are shown for the high Crude Oil prices (a)–(d) and for low Crude Oil prices (e)–(h) with MSTs on the left (a,c,e,g) and HTs on the right (b,d,f,h). Weekly (a,b,e,f) and monthly (c,d,g,h) frequencies are shown.

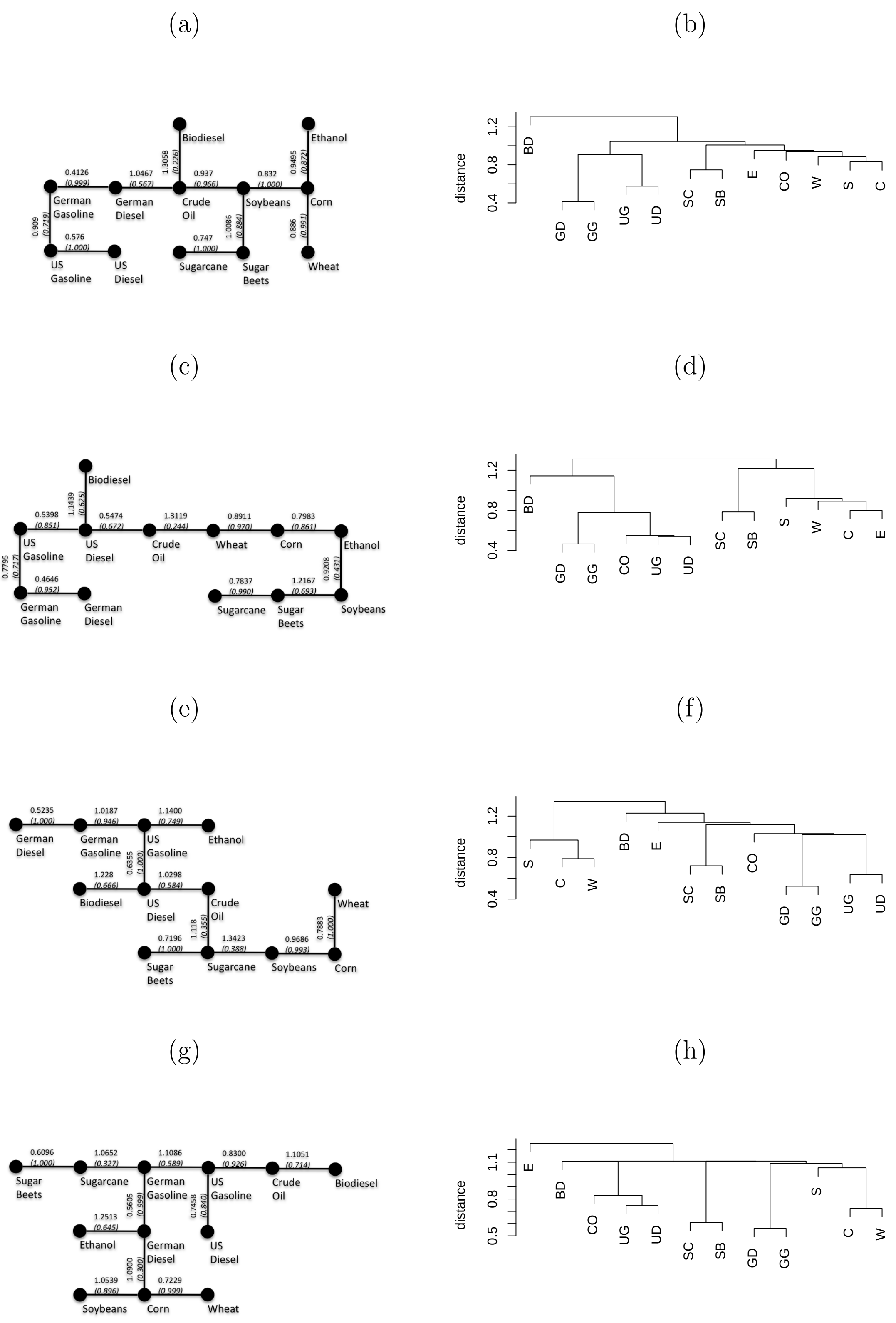

Figure 3: *Corn & Wheat.* Notation and ordering holds from Fig. 2.

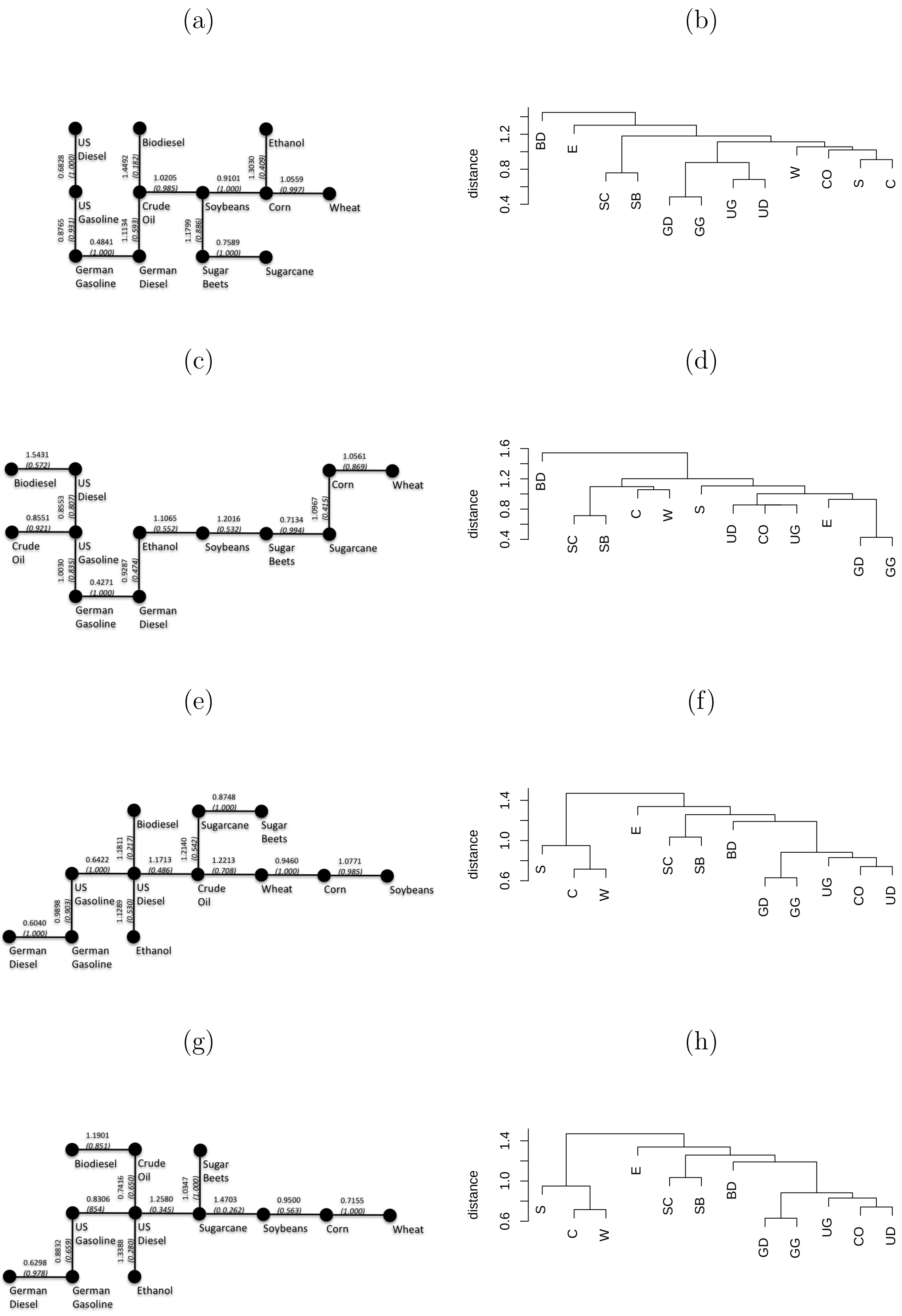

Figure 4: *Fuels.* Notation and ordering holds from Fig. 2.

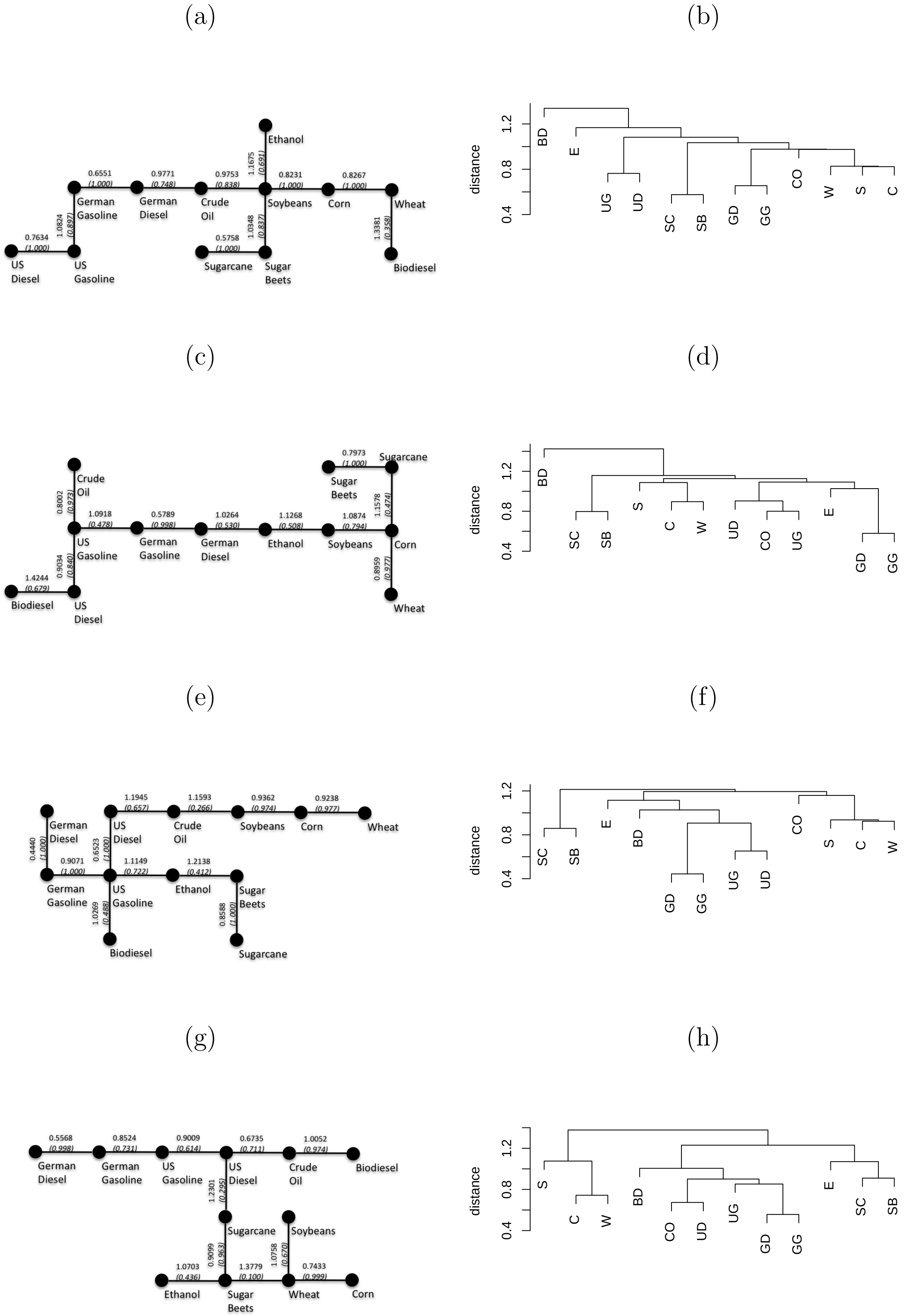

Figure 5: *Sugars.* Notation and ordering holds from Fig. 2.

Table 2: Stationarity tests – *ADF1 stands for ADF test with a constant, ADF2 stands for ADF test without a constant. KPSS test with one lag and a constant is used.*

Commodity	ADF1	p-value	ADF2	p-value	KPSS	p-value
Crude Oil	-13.3784	0.0000	-13.2848	0.0000	0.1157	> 0.1
Ethanol	-11.9235	0.0000	-11.9329	0.0000	0.0536	> 0.1
Corn	-13.3794	0.0000	-13.3284	0.0000	0.1249	> 0.1
Wheat	-12.7878	0.0000	-12.7795	0.0000	0.0916	> 0.1
Sugarcane	-14.3478	0.0000	-14.2195	0.0000	0.0923	> 0.1
Soybeans	-12.9502	0.0000	-12.9467	0.0000	0.1046	> 0.1
Sugar beets	-14.7365	0.0000	-14.6001	0.0000	0.0658	> 0.1
Biodiesel	-14.5312	0.0000	-14.3312	0.0000	0.1676	> 0.1
German diesel	-13.5802	0.0000	-13.5123	0.0000	0.1050	> 0.1
US diesel	-9.6001	0.0000	-9.4996	0.0000	0.2574	> 0.1
German gasoline	-12.6561	0.0000	-12.6276	0.0000	0.0693	> 0.1
US gasoline	-7.7812	0.0000	-7.7353	0.0000	0.1472	> 0.1

Table 3: Causality tests for ethanol and biodiesel. *(Note:* *, ** *and* *** *stand for a rejection of the null hypothesis "for* $X \to Y$, *X does not Granger-cause Y" and "zero aggregate effect", respectively for F and t-statistics, at* 10%, 5% *and* 1% *level of significance, respectively.)*

Impulse $\to$ response	F-statistic (causality) *4 weeks*	t-statistic (agg. effect) *4 weeks*	F-statistic (causality) *12 weeks*	t-statistic (agg. effect) *12 weeks*
$C \to E$	3.6163^{***}	2.9325^{***}	2.4339^{***}	2.77972^{***}
$S \to E$	1.2369	0.5019	0.7425	0.7973
$W \to E$	2.9285^{**}	2.4560^{**}	2.1188^{**}	1.6207
$SC \to E$	0.4305	0.6179	0.6776	1.2847
$SB \to E$	0.6940	1.0686	0.7577	2.1478^{**}
$E \to C$	0.2711	-0.5165	1.0000	-1.6541^{*}
$E \to S$	0.8580	-0.2626	0.8738	-1.0966
$E \to W$	0.8483	-1.1728	0.7865	-1.2830
$E \to SC$	1.3139	1.8467^{**}	0.9891	1.5894
$E \to SB$	0.6081	1.3459	0.6711	1.4448
$GD \to BD$	13.3305^{***}	6.5539^{***}	7.3931^{***}	3.9975^{***}
$CO \to BD$	10.3687^{***}	6.0689^{***}	6.9679^{***}	3.5331^{***}
$BD \to GD$	1.2895	-1.2499	1.2502	-1.3107
$BD \to CO$	2.2200^{*}	0.0397	1.1336	-1.4904