

Foods, fuels or finances: Which prices matter for biofuels?*

Ondrej Filip^a, Karel Janda^{a,b}, Ladislav Kristoufek^a, and David Zilberman^c

^aInstitute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, 110 00 Prague, Czech Republic

^bDepartment of Banking and Insurance, Faculty of Finance and Accounting, University of Economics, Namesti Winstona Churchilla 4, 130 67 Prague, Czech Republic

^cDepartment of Agricultural and Resource Economics, University of California, 207 Giannini Hall, Berkeley, CA 94720, USA

Abstract

We examine co-movements between biofuels and a wide range of commodities and assets in the US, Europe, and Brazil. We analyze a unique dataset of 33 commodities and relevant assets (between 2003 and 2016) which is unprecedented in the biofuels literature. We combine the minimum spanning trees correlation filtration to detect the most important connections of the broad analyzed system with continuous wavelet analysis which allows for studying dynamic connections between biofuels and relevant commodities and assets and their frequency characteristics as well. We confirm that for the Brazilian and US ethanol, their respective feedstock commodities lead the prices of biofuels, and not vice versa. This dynamics remains qualitatively unchanged when controlling for the influence of crude oil prices. As opposed to the Brazilian and US ethanol, the European biodiesel exhibits only moderate ties to its production factors. We show that financial factors do not significantly interact with biofuel prices.

Keywords: biofuels, prices, minimum spanning tree, wavelet coherence

JEL Codes: C22, C38, Q16, Q42

In various studies of price connections between biofuels and related commodities, a wide array of feedstock commodities has been used – corn, wheat, sugar cane, sugar beet, rapeseed, soybean, sunflower or palm oil [1]. However, it is possible and likely that interconnections between biofuels and other-than-feedstock commodities as well as other economic and financial variables play an important role as well [1–3]. Even though there

*Email addresses: ofilip38@gmail.com (Ondrej Filip), Karel-Janda@seznam.cz (Karel Janda), ladislav.kristoufek@fsv.cuni.cz (Ladislav Kristoufek), zilber11@berkeley.edu (David Zilberman). The research leading to these results was supported by Energy Biosciences Institute at University of California, Berkeley and by the European Union’s Horizon 2020 Research and Innovation Staff Exchange programme under the Marie Sklodowska-Curie grant agreement No 681228. The authors further acknowledge financial support from the Czech Science Foundation (grants number 16-00027S and 15-00036S). Karel Janda acknowledges research support provided during his long-term visits at McGill University, Toulouse School of Economics, New Economic School, Australian National University and University of California, Berkeley.

are several studies investigating the relationship between specific exchange rates and systems including oil and agricultural commodities [3–5], inclusions of stock indices, bond rates and interbank interest rates are rare. Also the existing studies considering financial variables did not explicitly include biofuels into the analyzed systems. An important motivation behind utilizing a wide portfolio of assets (regardless the theoretical existence or non-existence of links and connections among them) is the empirically found growing correlation among all types of assets with different fundamentals, which has become even more pronounced after the financial crises of the past decade, mainly due to financialization of commodities [6–8]. Investigating the dynamics and evolution of connections among biofuels and a wide range of other assets is thus novel and it is one of our main contributions.

Here we show that while the price movements of feedstock commodities are good predictors of biofuel prices for the US and Brazilian ethanol, this relationship is much weaker in Europe. We also show that the financial factors suggested in the literature on co-movement between oil and exchange rates [9–11], oil and stock market [12–14], and oil, agricultural commodities and exchange rates [3] do not significantly interact with biofuels prices. We find a clear market segmentation between conventional fuels and biofuels with respect to prices and that there is a spatial segmentation for different biofuels.

Results

We have collected a wide system of time series of weekly nominal prices for 33 biofuels-related commodities and assets over the period 2003-2016. This data was obtained from Bloomberg, Thomson Reuters Eikon, Centro de Estudos Avancados em Economica Aplicada (CEPEA), US Energy Information Administration (EIA), National Agency of Petroleum, Natural Gas and Biofuels – Agencia Nacional do Petroleo, Gas Natural e Biocombustiveis (ANP Brazil), US Federal Reserve, European Central Bank (ECB), and ECONSTATS databases. Such a wide system can easily become over-specified. To avoid this, we implement a two-step approach to the problem. We initially identify the most relevant connections in the studied system and then we focus only on these to uncover the specifics of the relationship between biofuels (namely the US and Brazilian ethanol, and European biodiesel) and the identified relevant factors.

In the first step of our approach, we examine the whole portfolio of assets suggested in the literature [1] as potentially related to biofuels and we apply the method of minimum spanning trees (MST) to identify the most relevant connections and co-movements between the factors during three major periods of the recent biofuels developments. While the MST approach allows for considering co-movements among essentially unlimited number of possible factors, it does not provide information on the direction of these co-movements. Therefore, as the second step of our analysis, we take the most relevant pairs identified by MST and we conduct an in-depth investigation of their mutual co-movements using an extended wavelet coherence framework. Technical details for both methodologies are given in the Methods section.

Minimum Spanning Trees

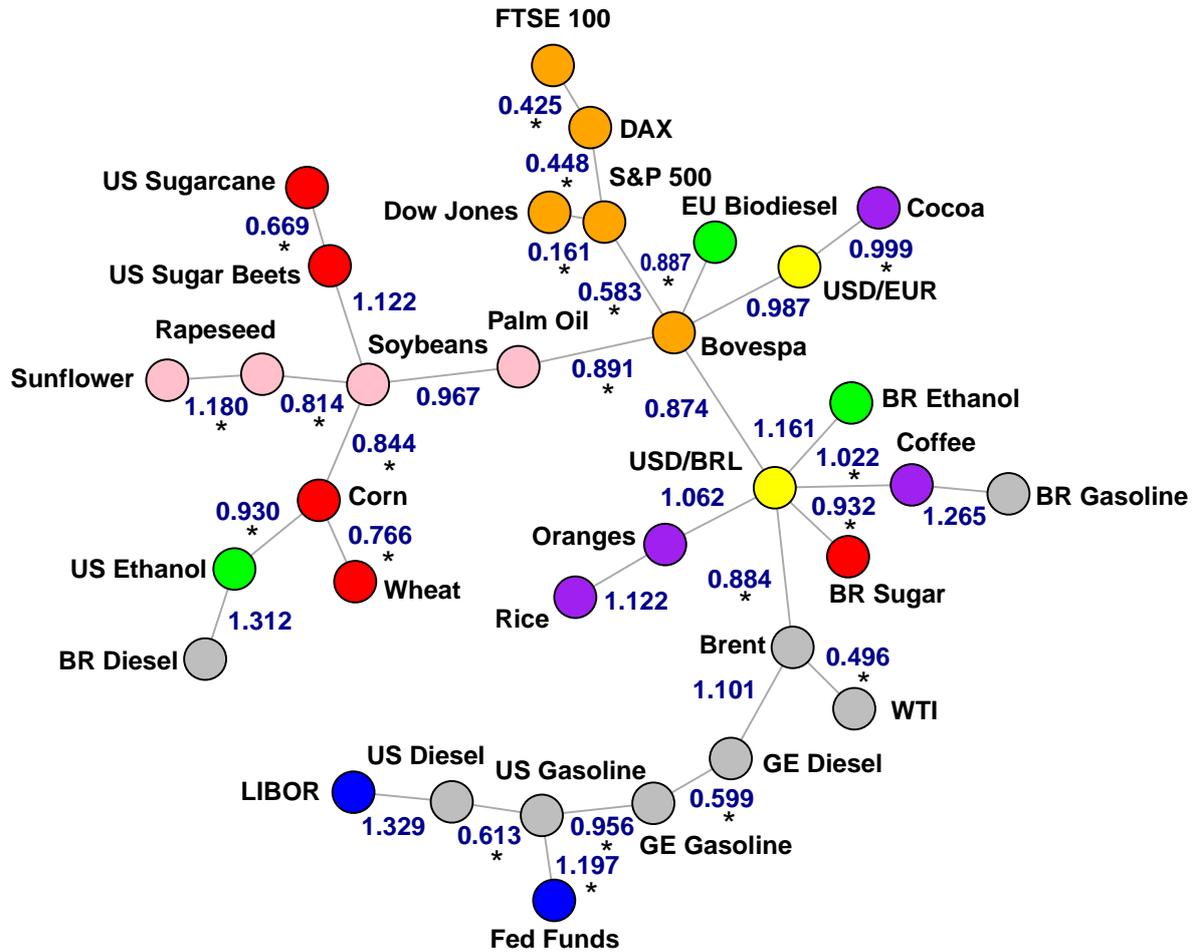
Inspecting Fig. 1, we observe that a core of the MST is formed by a compact group of the stock indices. It follows from the interconnected nature of stock markets that these indices are strongly correlated with stable mutual links. We see that not only stock indices but also a number of other items gather according to their type. In broader terms, there seems to be a group of agricultural commodities — a food branch — and a fuel branch of the MST at the opposite sides of the network. On the one hand, a cluster of the vegetable oils together with the ethanol feedstock commodities constitute the food part of the tree. On the other hand, the fossil fuels and biofuels form the fuel branch. While no other biofuel is connected to its feedstock, Brazilian ethanol makes a notable exception. Its robust link to Brazilian sugar is evident already during this first studied period (2003 – 2008).

The US ethanol and even biodiesel have stable links to their US fossil substitutes and crude oil. This close relationship between the US ethanol and US gasoline prices indicates that the US policies supporting ethanol use were still in their early stage and their influence on prices of US ethanol feedstocks was not a binding factor for pricing of ethanol. Instead, the US ethanol prices were determined by oil price, which experienced stable strong growth during the whole 2003-2008 period. During this early period, the ethanol production capacity was limited and the ethanol refinery industry was expanding rapidly. The expansion was driven by high ethanol prices that in addition to increasing crude oil prices reflected the blenders' tax credit, the need for oxygenated fuel additives replacing MTBE additives, which had been banned in some of the US states starting from 2002, and government blending mandates operational since 2005.

As opposed to the exchange rates, which bridge the stock market cluster with the fuel and food parts, interest rates do not seem to interact with a rest of a system a lot. Brazilian retail fuels are not integrated into the fuel branch, they stand at the edge of the network being only linked to LIBOR. The isolated position of the Brazilian fossil fuels is implied by a specific setting of the national fuel market in Brazil [15]. Due to a decisive influence of Petrobras on local fuel prices, Brazilian fossil fuels do not necessarily follow the global markets' development. As we will see during the whole studied period, they do not usually integrate into the fossil fuel cluster. We should not forget about the four purely food commodities which cannot be used to produce biofuels. These items do not form any cluster and are only individually connected to different nodes of the network.

Our first period covered the era of rising food prices and an increasing global significance of biofuels preceding the first world food crisis covered by our sample. We now continue investigating our network's development in a changed market environment. After several years of accelerating agricultural and energy prices, these slumped quickly during the second half of 2008 hand-in-hand with the global economic crisis. After the bottom was reached in 2009, both energy and agricultural prices started a new rally up to a new peak in February 2011. During this period, we expect to see signs of an established biofuel production not only in Brazil but also in the US and Europe. Fig. 2 delivers a minimum spanning tree for this period. In this period, biodiesel gets closer to the vegetable oils cluster and the Brazilian ethanol remains close to the Brazilian sugar and the USD/BRL exchange rate. Compared to the previous period, a major change occurs for the US ethanol which moved to the food branch, connecting to corn. The US ethanol–

Figure 2: Minimum spanning tree (2008 – 2011)



Minimum spanning tree for the second period (2008 – 2011) is shown here. Colors represent types of commodities – biofuels (green), ethanol feedstock (red), biodiesel feedstock (pink), other agricultural commodities (purple), fossil fuels (gray), stock indices (orange), exchange rates (yellow) and interest rates (blue). Distances are shown and length of links between commodities is proportional to the distances. Stability of links is tested using the bootstrapping procedure and the stable links are marked by an asterisk. Again, most commodity types cluster nicely together but there are some notable differences compared to the first analyzed period. Brazilian ethanol is still close to sugar but it has moved closer to the exchange rate between USD and BRL and Brazilian gasoline. EU biodiesel has moved closer to its producing factors and the US ethanol is now closely connected to corn and wheat.

ducers and the capacity expansion stopped. This was also connected with reaching the so-called blending wall (the 10 percent limit on the ethanol content in gasoline).

Inspecting the minimum spanning tree in Fig. 3, we notice the central position of fossil fuels cluster which clearly separates the US and EU biofuels on one side and Brazilian ethanol on the other side. Exceptionally, both Brazilian fossil fuels get integrated into

interconnections between palm oil, rapeseed, and soybeans with biodiesel being attached to both sunflower and palm oil at the same time as keeping close connection with fossil fuels through Brent crude oil.

Wavelet Analysis

The minimum spanning tree analysis in the previous sections has shown that there is quite noticeable dynamics of the evolution of the MST structure over time. Given the static nature of each MST structure depending on the chosen time window for that particular MST, there is a clear advantage in using continuous wavelet framework capturing the co-movement among the prices of chosen commodities at each time point and at different scales (frequencies) without a need to specify any particular time window or choosing a particular fixed frequency. In other words, the continuous wavelet framework allows us to study the time evolution of correlation between series without any parametrization due to moving window length or possible structural breaks detection. Along the same line, the possibility of studying the relationship at different scales uncovers more details and provides more information about its dynamics, which would otherwise remain hidden or averaged out over all scales (e.g. using standard correlation or regression frameworks). Having identified several important links between biofuels and their feedstock, we approach these pairwise connections separately.

The output for each studied biofuel–feedstock pair is presented in form of two charts. While the horizontal axes show time (in years), there is also the frequency scale (in days) on the vertical axes. Coherence is indicated by color according to a spectrum shown at the right edge. Pale colored corner areas are not of reliable interpretation. They result from artificially adding zeros to the beginning and to the end of analyzed series compensating for wavelet lengths. A central bright colored area delivers reliable results. Furthermore, the regions with statistically significant coherence are bounded with a thick black curve.

In the left panel, we present the squared wavelet coherence between biofuel and a particular relevant feedstock commodity indicated by a combination of MST results and other considerations. Since there is no negative wavelet coherence, phase difference between the series is indicated by directed arrows. Simply put, the arrows indicate a direction of the relationship. Rightward pointing arrows mean that biofuel is positively correlated with that particular feedstock while leftward pointing arrows indicate a negative relationship. If the arrows point straight down, biofuel leads the price of feedstock by $\pi/2$. On the contrary, the upward pointing arrows imply that price of biofuel is led by feedstock.

In the right panel, we show the partial wavelet squared coherence. In the food–fuel system, crude oil plays a role of an important price driver affecting both fuel and food parts of the commodity system. In particular during the 2003-2008 period, both EU biodiesel and US ethanol were part of fuels branch of MST with both Brent and WTI crude oil being close to these biofuels. This close relationship diminished during crisis period 2008-2011. However, in the following 2011-2016 period, biodiesel price became again closely related to Brent crude oil price while keeping close to the well-defined MST cluster of its oilseed feedstocks. Over all three analyzed periods, the Brazilian ethanol was always consistently connected with the main fossil fuels MST cluster (including crude oil) through financial factors (exchange rates and stock indices). Therefore, the wavelet coherence output (left panel) is always supplemented by the partial wavelet coherence

chart (right panel) where crude oil is controlled for. The increase of the price co-movement between oil and food commodities with the onset of the global food crises has been previously indicated [16–18] and it is a central question of both earlier [19–21] and most recent [22–24] literature dealing with the role of biofuels as an important link in the development of food price dynamics.

Brazilian Ethanol

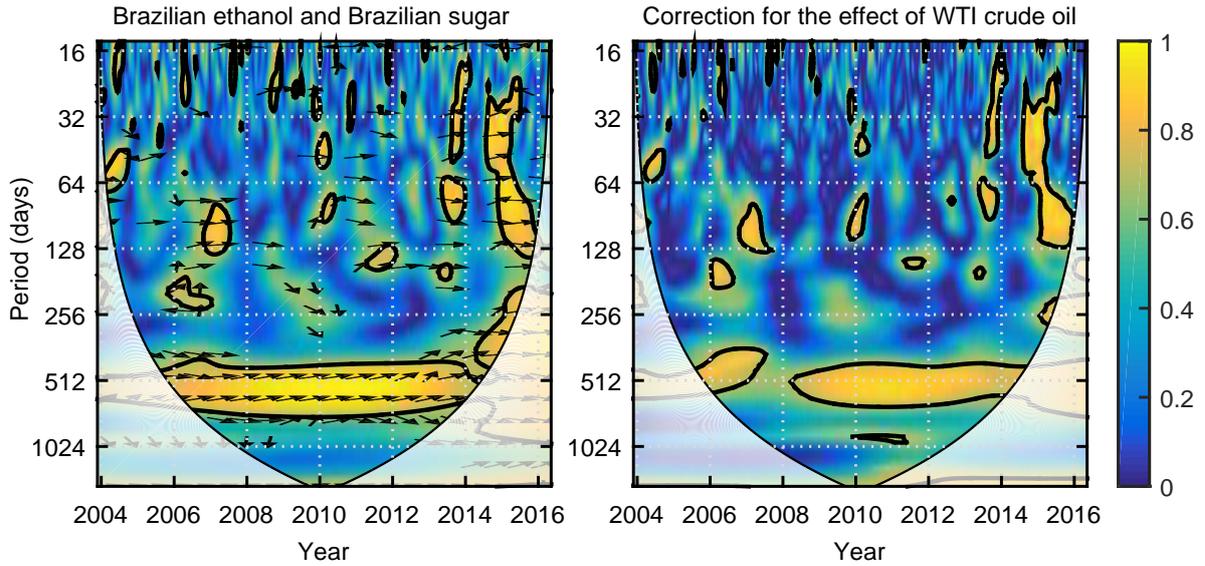
As documented by our previous MST analyses, Brazilian ethanol price is directly linked to Brazilian sugar price, which may be taken as a best available market representation of the value of sugarcane serving as ethanol feedstock. Moreover, their connection seems to be quite strong and stable since it appears in majority of the constructed MSTs. Hence, we are especially keen on exploring Brazilian ethanol (BR ethanol) correlation with local sugar (BR sugar) price in more detail.

The left panel of Fig. 4 shows a strong relationship apparent for scales approximately between 500 and 700 days, i.e. roughly between 1.5 and 2 years. Thus, we observe a long-term relationship between Brazilian ethanol and Brazilian sugar that is remarkably stable in time. Apart from this main relationship, we identify only several minor coherence islands associated with relatively quick price interactions in the short term. For 2015, we observe an intensive co-movement of ethanol and sugar on essentially all observed frequencies. The phase arrows in the main significant region point to the right and upwards indicating a positive correlation between ethanol and sugar prices with sugar leading ethanol.

The right panel of Fig. 4 delivers the output of partial wavelet coherence when the crude oil influence has been controlled for. Apparently, crude oil consumed only a small portion of the overall correlation. The stable long-term relationship between Brazilian sugar and ethanol reported in the left panel is distracted for a short period around 2008 after controlling for the crude oil influence. This time period overlaps with historical heights of crude oil prices when a barrel of crude oil rocketed over \$140 in July 2008. This suggests that crude oil has a stronger effect on ethanol when the prices are high. Such claim is well in line with the results detecting different non-linear relationships between crude oil, biofuels, and their feedstocks, specifically with respect to high crude oil prices [19, 25–29]. Our results describe the dynamics in more detail. Specifically, only the long-term relationship is formed for the crude oil highs but there are no emerging short-term correlations as a result of the crude oil spikes. The relationship between Brazilian ethanol and sugar thus remains remarkably stable in time, which has been reported in the MST analysis as the connection between the Brazilian sugar and ethanol having been found strong and reliable during all three analyzed periods.

Wavelet coherence between Brazilian ethanol and Brazilian sugar is noticeably less sensitive to the correction for the effect of world crude oil price, as compared to US ethanol and European biodiesel, which are covered in the two following subsections. This reflects the insulation of Brazilian fuel markets from the world crude oil prices fluctuations provided by the domestic wholesale oil price cap effectively enforced by the state-owned oil producer, Petrobras, which sets a reference oil price for domestic refining [15]. Given that this reference price was usually set below the world market price, this implicit subsidy to gasoline provides a negative incentive for biofuel production which clashes with Brazilian

Figure 4: Wavelet coherence for Brazilian ethanol and Brazilian sugar.



Both wavelet squared coherence (*left*) and partial wavelet coherence (*right*) correcting for the effect of WTI crude oil are shown. Coherence color spectrum is shown at the right edge, i.e. the hotter the color the higher the coherence. Regions with wavelet coherence statistically significant at 95% level are highlighted by a thick black curve. The significance is tested against the null hypothesis of a red noise and the procedure is based on 1,000 simulations. We find a stable significant region at the scale of around 500 days, i.e. approximately 2 trading years, which holds even after controlling for the crude oil influence. Orientation of the phase shift arrows suggests that the sugar series lead the ethanol series, i.e. the feedstock leads the biofuel and not vice versa.

biofuel mandate and fuel tax/tax credit incentivizing biofuel.

US Ethanol

Our results (Fig. 5) show that the relationship between the US ethanol and corn (which accounts for 90% of the US ethanol) consists of two strong dependencies of different kinds. We find significant coherence areas associated with both short-term and long-term horizons. The long-term relationship approximately at the level of 500 days (almost 1.5 years) has been steadily present since the period following the food crisis of 2008. Its rightwards pointing phase arrows tell us that the US ethanol has been positively correlated with corn throughout the second half of the studied time frame. Compared to the Brazilian case, the feedstock leadership is not as evident.

The other type of dependency is a collection of short-term price interactions which are, however, much stronger compared to the Brazilian case. These time-limited episodes are associated with very high corn prices in summer 2008, first half of 2011 and second half of 2012. The first two cases coincide with food price crises of 2008 and 2011. The 2012 strong co-movement of ethanol and corn prices is associated with the 2012 US drought which was the most severe drought in the US since 1956. Our results support the findings

of [30] and [31], who focus on the interplay of (weather influenced) grain commodities storage and biofuels policies in the determination of prices of grains and ethanol.

Altogether, we find a stable long term relationship accompanied by several short-term episodes associated with high corn prices especially between 2010 and 2013. This supports the findings of the MST analysis which identified a strong and stable connection between the US ethanol and corn mainly during the later two periods, i.e. after 2008. Throughout the last decade, the relationship has always been positive with corn slightly leading the price of ethanol. When the influence of crude oil is controlled for, we apparently lose a part of correlation. Especially the long term relationship between ethanol and corn gets somewhat reduced for the period of global food crisis 2008-2011. However, other qualitative results do not get affected.

Comparison of price dynamics revealed by both MST and wavelet coherence results for the US and Brazilian ethanol shows that in the US the Renewable Fuel Standards (RFS) had a stronger impact than the corresponding Brazilian policies. This is because in Brazil other factors like already mentioned upper bound on wholesale price of gasoline, different taxations for biofuel and fossil fuel or subsidization of diesel [15] together with a long term history of the world's most developed biofuel market implied nonexistence of pronounced RFS caused regime switch which we documented on the US data.

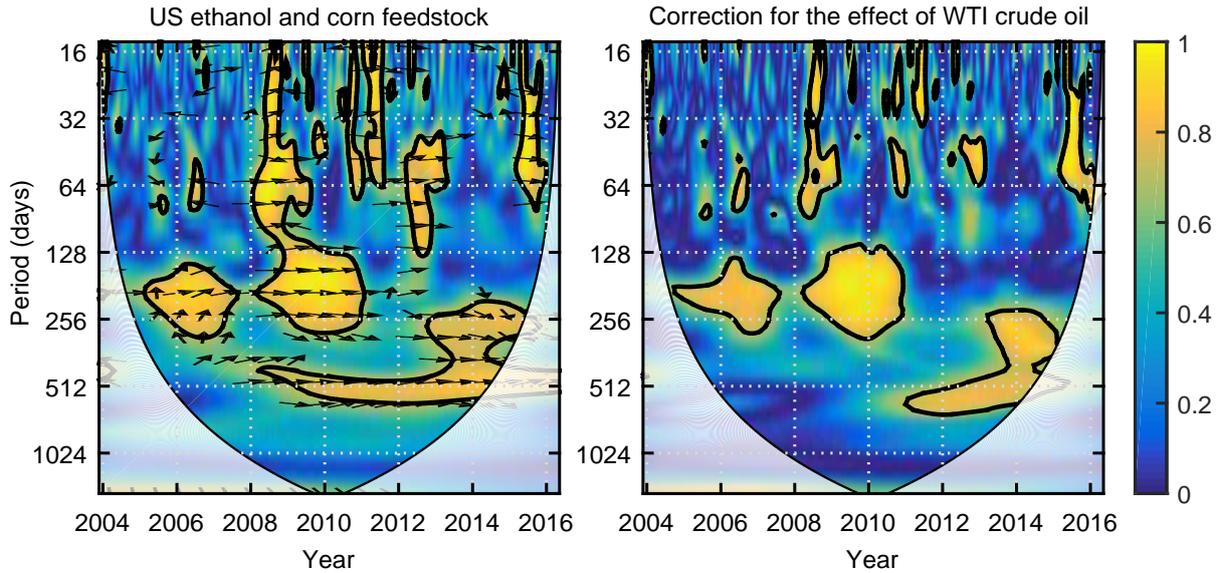
European Biodiesel

The minimum spanning tree analyses have indicated that biodiesel exhibited a price behavior different from both the US and Brazilian ethanol – its position in the network evolved vividly over the analyzed periods. Now we focus on the relationship between EU biodiesel and rapeseed as its main European producing factor and possible influence of Brent crude oil (Fig. 6).

In the left panel of the figure, we find a strong and stable relationship between biodiesel and rapeseed. The relationship is positive and rapeseed leads the dynamics of biodiesel according to the phase arrows direction. The connection is visible for high scales up to approximately 1000 days, i.e. 4 years, which could be a reason why such connection has not been detected by the MST analysis. The range of significant scales shrinks with time and after 2014, the connection is gone apart from few small islands at lower scales. However, this relationship between biodiesel and its main feedstock is mainly due to the correlation of the constituents and crude oil. Specifically, the right panel of Fig. 6 uncovers that after controlling for the effect of Brent oil, the connection between rapeseed and biodiesel vanishes practically completely and it has been found due to an indirect correlation, i.e. changes in Brent crude oil cause price changes in both rapeseed and EU biodiesel but there is no direct long-lasting causality coming from rapeseed towards biodiesel. Similar results of low correlation between prices of biodiesel and prices of its feedstock are also characteristic for other feedstock besides rapeseed which we covered in our exploratory analysis (these results are available upon request) and which we do not report in this paper. Since the European retail fuel prices are higher than those in the US (because of low US taxes) and Brazil (because of lower wholesale crude oil price in Brazil) with fuel/food price ratios being higher in Europe than in the US, the European biodiesel prices are aligned with fossil fuels rather than with foods.

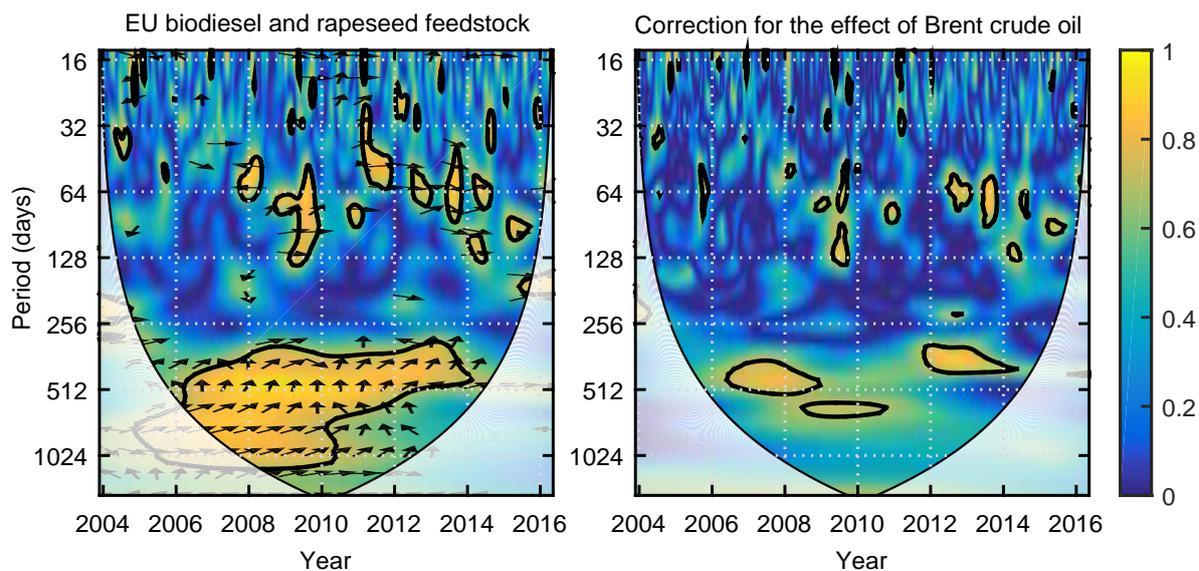
Nonetheless, there are certain signs of a time-limited positive correlation between

Figure 5: Wavelet coherence for US ethanol and corn feedstock



Both wavelet squared coherence (*left*) and partial wavelet coherence (*right*) correcting for the effect of WTI crude oil are shown. Coherence color spectrum is shown at the right edge, i.e. the hotter the color the higher the coherence. Regions with wavelet coherence statistically significant at 95% level are highlighted by a thick black curve. The significance is tested against the null hypothesis of a red noise and the procedure is based on 1,000 simulations. Similarly to the case of Brazil, we find a relatively stable significant region at the scale of around 500 days, i.e. approximately 2 trading years. However, it is more weakened by controlling for the crude oil effect and, importantly, the strong relationship emerges only after the first food crisis, i.e. after 2008. For this period immediately after 2008, the positive correlation is evident for practically all scales. After 2011, the connection weakens for lower scales. Orientation of the phase shift arrows suggests that corn leads ethanol for most time-scale points. However, such leadership is again weaker than for the Brazilian case.

Figure 6: Wavelet coherence for biodiesel and rapeseed feedstock



Both wavelet squared coherence (*left*) and partial wavelet coherence (*right*) correcting for the effect of WTI crude oil are shown. Coherence color spectrum is shown at the right edge, i.e. the hotter the color the higher the coherence. Regions with wavelet coherence statistically significant at 95% level are highlighted by a thick black curve. The significance is tested against the null hypothesis of a red noise and the procedure is based on 1,000 simulations. We find a stable significant region at the scales above 300 days, i.e. above a trading year. Orientation of the phase shift arrows suggests that the feedstock commodities lead the biodiesel prices. However, this strong relationship is practically completely gone after controlling for the Brent oil prices. This suggests that the emergent relationship between biodiesel and its feedstock is mainly due to indirect correlation with crude oil.

biodiesel and its feedstock. During these episodes, feedstock led the price of biodiesel. Compared to the results obtained for the Brazilian and US ethanol, we argue that the price of European biodiesel is very weakly connected with prices of individual feedstock commodities. Its price is primarily determined by a mix of factors including fossil fuels prices and the price behavior of the whole oilseed cluster rather than only price of a single dominant feedstock. In this respect, the European biodiesel market substantially differs from the analyzed ethanol markets. The difference in the terms of close connection of EU biodiesel with fossil fuels, as compared to Brazil and US ethanol close connection to feedstock, may be explained by a key role played by government policies in the determination of biofuel prices. In the case of Brazil and the US, the fossil fuel price is not determining ethanol prices because the well defined mandates (particularly strong in Brazil with the ethanol content of motor vehicle fuels being mandated as high as 27 percent recently) tend to isolate the fossil fuel markets and ethanol markets from each other. This was not the case in the EU where especially during the initial period before the global food crisis the biodiesel tax exceptions played important role. Following the economic mechanism described in [22, 32], the continuing importance of co-movement of EU biodiesel and fossil fuels prices is also supported by nonexistence of EU-wide binding mandates as opposed to the situation in the US and Brazil.

Having analyzed the EU biofuel industry in the context of its main regulatory drivers, we conclude that European biodiesel plays a different role than the US and Brazilian ethanols do in their domestic markets. The difference has been demonstrated by the results of the MST and wavelet analyses.

Discussion

Our paper is an integral part of the wide literature on biofuels-related price transmission. The most comprehensive recent treatment and literature overview of this topic is provided in [22] which is complemented by a series of related papers [32–36] focused mainly on biofuels policies and prices. Our analysis has focused on the world’s major biofuel markets – Brazil, the US, and the EU. Our study covers 83% of the global ethanol production and about 45% of the biodiesel production. We have studied the relationships between ethanol, biodiesel, associated agricultural commodities, crude oil, relevant fossil fuels, and a group of financial assets. In summary, we have succeeded in confirming our initial hypotheses. First, we have described an interconnected system of the biofuel-related commodities. Moreover, we have commented on its evolution over the course of a fourteen-year period. Second, we have documented a phase shift that initially occurred between the mature Brazilian and belated US/European biofuel industries. Third, we have demonstrated a positive price co-movement of ethanol and its respective production factors. We have further showed that this relationship is stable in time with feedstock leading the price of ethanol. Fourth, we have explained that the price of biodiesel did not depend on a single feedstock commodity. Biodiesel weakly interacted with several crops through more random price adjustments and with fossil fuels. Thus, the European biofuel industry substantially differs from both the Brazilian and the US establishments. Finally, the decoupling of US ethanol from the fuels cluster as observed in our MST analysis illustrates an important role played by government policies in influencing bio-

fuels prices. In agreement with biofuels policy theoretical models of [22, 32], the firm establishment of US ethanol mandates and suppression of tax incentives led to closer ties between US ethanol and its feedstock prices. Similar policy argument of gradual shift from tax incentives to mandates also explains the price co-movements of EU biodiesel. The results of our MST and wavelet coherence therefore show a significant impact of government policy decisions (in our case the choice of biofuels support instruments) on market transmitted price behavior.

The main contribution of this paper lies in its innovative and comprehensive approach. Employed methods make as few *ex ante* assumptions as possible. In particular, the wavelet coherence methodology represents a widely applicable non-parametric toolbox without assuming any specific underlying model. The method is simply built on a superposition of time and scale bounded effects. Our findings contribute to the current biofuel policy discussion. Specifically, we stress the difference between ethanol and biodiesel production processes. Eventually, we shed some light on the biofuel–feedstock connections on the leading global markets.

Methods

Minimum Spanning Trees

The minimum spanning trees (MST) provide a metric to measure the interconnections within a group of commodities and assets. The construction of MST builds on a sample pairwise Pearson correlation coefficient $\hat{\rho}_{ij}$ between assets i and j . The correlation coefficients are transformed into a distance measure [37]

$$d_{ij} = \sqrt{2(1 - \hat{\rho}_{ij})}. \quad (1)$$

Unlike $\hat{\rho}_{ij}$, the distance measure d_{ij} is a true distance measure fulfilling the axioms of Euclidean metric and it ranges between 0 and 2. Specifically, the corner cases of the correlation coefficient can be translated into the distance measure language in the following manner:

$$d_{ij} = \begin{cases} 0 & \text{perfect positive correlation;} \\ \sqrt{2} & \text{no correlation;} \\ 2 & \text{perfect negative correlation.} \end{cases}$$

Since each correlation coefficient $\hat{\rho}_{ij}$ is transformed into a distance measure d_{ij} , the correlation matrix \mathbb{C} is transformed into a distance matrix \mathbb{D} :

$$\mathbb{C} = \begin{pmatrix} \hat{\rho}_{11} & \hat{\rho}_{12} & \dots & \hat{\rho}_{1n} \\ \hat{\rho}_{21} & \hat{\rho}_{22} & \dots & \hat{\rho}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{n1} & \hat{\rho}_{n2} & \dots & \hat{\rho}_{nn} \end{pmatrix} \Rightarrow \mathbb{D} = \begin{pmatrix} 0 & d_{12} & \dots & d_{1n} \\ d_{21} & 0 & \dots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & 0 \end{pmatrix} \quad (2)$$

Using the values of \mathbb{D} , we find the most important connections within a group of commodities and assets according to the MST concept using the Kruskal's algorithm [38]. In short, the tree of the most important connections of the system are formed by systematically eliminating the weakest links (the longest distances) between nodes (commodities and assets) as long as the tree is not torn between more trees, i.e. as long as it is possible to connect any two nodes using the remaining links. The number of links decreases from $k(k-1)/2$ for the correlation matrix to $k-1$ for the final minimum spanning tree, where k is the number of variables. Such correlation matrix filtration has proved its worth in various applications across disciplines [39–42].

A potential issue of the MST analysis lays in a possible link instability, i.e. whether the detected relevant link is in fact relevant or is present due to statistical noise. This is usually an issue for systems with overall weak connections. In order to assess stability (importance) of individual links, we employ the bootstrapping technique proposed by [43]. In the procedure, the time series are re-sampled with repetition from the original series while the variable of re-sampling is time, i.e. we re-sample from the time index and reconstruct the analyzed series according to the new time structure. This way, the pairwise correlation core is not distorted. The MST procedure is then applied on the bootstrapped series and the relevant links are recorded. This is repeated 1000 times. Resulting bootstrap values are obtained for each link in form of b_{ij} which is defined as a ratio between the number of occurrences by the total of bootstrapped realizations so that $b_{ij} \in [0; 1]$. The links with $b_{ij} > 0.5$ are denoted in our MST charts with an asterisk.

Wavelet Coherence

The most important connections in the analyzed system of commodities and assets identified through MST are further investigated in more detail using the wavelet coherence framework, which is a significant generalization of the simple correlation analysis as it allows for examination of the time evolution of the relationship between series as well as its scale structure, i.e. whether the connection is important from different temporal perspectives.

Wavelet analysis decomposes a time series into several components according to their time and scale properties. Generally, a signal (time series) may be composed of individual waves cycling at different speed, i.e. with different frequencies. Individual components of the signal get separated in the frequency domain. A central feature of wavelet analysis is that it captures both time and frequency characteristics, i.e. it decomposes the series in both domains.

A wavelet $\psi_{u,s}(t)$ is a real or complex-valued function given as

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right), \quad (3)$$

with a scale parameter s and a location parameter u . Under certain conditions [44], the original series $\{x_t\}$ can be fully reconstructed from its wavelet transform $W_x(u, s)$ defined as

$$W_x(u, s) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt, \quad (4)$$

where $*$ represents a complex conjugate operator preventing an information loss during the transformation. The degree of similarity between the wavelet shape and $\{x_t\}$ is measured by the integral above. Frequently, multivariate economic applications of wavelet analysis use the Morlet wavelet which, as a complex wavelet, enables studying multivariate relationships between series. The Morlet wavelet is specified as $\psi^M(t) = \frac{1}{\pi^{1/4}} e^{i\omega_0 t} e^{-t^2/2}$, where ω_0 denotes the central frequency of the wavelet. We set $\omega_0 = 6$ as it provides an optimal balance between time and frequency components [45–48]. As an example of a different wavelet used in energy economics we may consider Haar a Trouse wavelet [13].

The wavelet framework can be generalized into a bivariate setting which is essential for studying connections between various series. In the bivariate setting, the cross wavelet spectrum is given by

$$W_{xy}(u, s) = W_x(u, s)W_y^*(u, s), \quad (5)$$

where $W_x(u, s)$ stands for the continuous wavelet transform of series $\{x_t\}$ and $W_y^*(u, s)$ marks a complex conjugate of the continuous wavelet transform [49]. As the cross wavelet spectrum is generally complex, the cross wavelet power is given by $|W_{xy}(u, s)|$. It is usually understood to be a measure of local covariance between two series at a given frequency. Nonetheless, we cannot easily assess the strength of the detected co-movement as the cross-wavelet power is not bounded (in the same logic as standard covariance). To overcome such limitation, the squared wavelet coherence is introduced as

$$R_{xy}^2(u, s) = \frac{|S(\frac{1}{s}W_{xy}(u, s))|^2}{S(\frac{1}{s}|W_x(u, s)|^2)S(\frac{1}{s}|W_y(u, s)|^2)}, \quad (6)$$

with S being a smoothing operator [50]. By definition, the value of the squared coherence varies between 0 and 1. Moreover, the squared wavelet coherence corresponds to the usual squared correlation coefficient for specific time and frequency. As the cross-wavelet spectrum translates into the squared coherence, the information about the direction of the relationship is lost. However, this information can be recovered from its phase difference specified as

$$\varphi_{xy}(u, s) = \tan^{-1} \left(\frac{\Im [S(\frac{1}{s}W_{xy}(u, s))]}{\Re [S(\frac{1}{s}W_{xy}(u, s))]} \right), \quad (7)$$

with \Re and \Im representing the real and imaginary part operators, respectively. For technical details, please refer to [51].

Wavelet coherence is limited by the same technical constraint as usual correlation – it may suffer from the omitted variable bias as it does not control for a possible influence of other variables [52]. Thus, we may observe a (seemingly) high coherence between two variables while the observed relationship can in fact be caused by their mutual ties to another variable. To overcome this issue, we follow [52] in using the partial wavelet squared coherence, an analogy of the partial correlation, defined as

$$RP_{y,x_1,x_2}^2 = \frac{|R_{yx_1} - R_{yx_2}R_{yx_1}^*|^2}{(1 - R_{yx_2}^2)(1 - R_{x_2x_1}^2)}. \quad (8)$$

Partial wavelet coherence evaluates the relationship between $\{y\}$ and $\{x_1\}$ while controlling for the effect of $\{x_2\}$, see [53] for details.

There are two main sources of uncertainty present when translating the time series into the time-frequency domain of wavelet coherence. First, a high level of wavelet coherence at specific time and scale point does not guarantee strong relationship between series at the given time-scale point. Statistical significance is inferred from Monte Carlo simulations. Specifically, the null hypothesis set here is a red noise, i.e. an autoregressive process of the first order. Such process has stable short-term memory autocorrelated dynamics in both time and frequency domain. The 95% significance bands represented by a thick black curve in the wavelet coherence contours are based on 1,000 simulations. We thus follow the standard procedure set in the applied wavelet coherence literature [50, 51]. Second, wavelets are being rescaled for specific scales during the procedure and projected onto the time series. Such rescaling (stretching) meets a limit for higher scales as the sample size is limited. Therefore, the beginnings and the ends of the analyzed time series cannot be properly represented for high scales and the time series needs to be padded by zeros to provide sufficient number of observations. This makes these regions less reliable than the ones based on the actual data. The reliable regions of the time-scale plane of the wavelet coherence are separated from the less reliable ones by the so-called cone of influence which splits the plane into two – the less reliable with pale colors and the reliable with full colors. The wavelet coherences shown in full colors can be thus interpreted and commented on without issues [51].

In the same way as for the minimum spanning tree approach, an outcome of the wavelet analysis is best understood from its graphical representation, which is discussed and described in detail in the Results section. The computational process of our wavelet analysis was done in MATLAB R2014b using packages by Aslak Grinsted and E. K. W. Ng and T. W. Kwok.

Contributions

Authors (O.F., K.J., L.K. and D.Z.) have contributed to the manuscript equally.

Competing interests

The authors declare no competing financial interests.

References

- [1] Teresa Serra and David Zilberman. Biofuel-related price transmission literature: A review. *Energy Economics*, 37(0):141 – 151, 2013.
- [2] Özge Savaşçın. The dynamics of commodity prices: A clustering approach. *University of North Carolina, Chapel Hill*, 2011.
- [3] Anthony N. Rezitis. The relationship between agricultural commodity prices, crude oil prices and us dollar exchange rates: a panel VAR approach and causality analysis. *International Review of Applied Economics*, 29(3):403–434, 2015.

- [4] S. Nazlioglu and U. Soytas. Oil price, agricultural commodity prices, and the dollar: a panel cointegration and causality analysis. *Energy Economics*, 34:1098–1104, 2011.
- [5] S. Nazlioglu and U. Soytas. World oil prices and agricultural commodity prices: evidence from an emerging market. *Energy Economics*, 33:488–496, 2011.
- [6] K. Tang and W. Xiong. Index investment and the financialization of commodities. *Financial Analysts Journal*, 68(6):54–74, 2012.
- [7] B. Büyüksahin and M.A. Robe. Speculators, commodities and cross-market linkages. *Journal of International Money and Finance*, 42:38–70, 2014.
- [8] K. Avdulaj and J. Barunik. Are benefits from oil-stocks diversification gone? New evidence from a dynamic copula and high frequency data. *Energy Economics*, 51:31–44, 2015.
- [9] Juan C. Reboredo. Modelling oil price and exchange rate co-movements. *Journal of Policy Modeling*, 34(3):419–440, May-June 2012.
- [10] Juan C. Reboredo and Miguel A. Rivera-Castro. A wavelet decomposition approach to crude oil price and exchange rate dependence. *Economic Modelling*, 32:42–57, May 2013.
- [11] Juan C. Reboredo, Miguel A. Rivera-Castro, and Gilney F. Zebende. Oil and US dollar exchange rate dependence: A detrended cross-correlation approach. *Energy Economics*, 42:132–139, March 2014.
- [12] Andrea Bastianin and Matteo Manera. How does stock market volatility react to oil shocks? Working Paper 2015-09, Department of Economics, Management and Quantitative Methods at Università degli Studi di Milano, June 2015.
- [13] Rania Jammazi and Juan C. Reboredo. Dependence and risk management in oil and stock markets. A wavelet-copula analysis. Unpublished, 2015.
- [14] Juan C. Reboredo and Miguel A. Rivera-Castro. Wavelet-based evidence of the impact of oil prices on stock returns. *International Journal of Economics and Finance*, 29:145–176, January 2014.
- [15] Madhu Khanna, Hector M. Nunez, and David Zilberman. Who pays and who gains from fuel policies in Brazil? *Energy Economics*, 54:133–143, February 2016.
- [16] Juan C. Reboredo. Do food and oil prices co-move? *Energy Policy*, 49:456–467, October 2012.
- [17] 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.
- [18] Jan Pokrivcak and Miroslava Rajcaniova. Crude oil price variability and its impact on ethanol prices. *Agricultural Economics – Czech*, 57(8):394–403, August 2011.

- [19] 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.
- [20] Pavel Ciaian and d’Artis Kancs. Food, energy and environment: Is bioenergy the missing link? *Food Policy*, 36(5):571–580, October 2011.
- [21] Andrea Bastianin, Marzio Galeotti, and Matteo Manera. Biofuels and food prices: Searching for the causal link. Working Paper 239, University of Milano-Bicocca, Department of Economics, March 2013.
- [22] Harry de Gorter, Dusan Drabik, and David R. Just. *The Economics of Biofuel Policies. Impacts on Price Volatility in Grain and Oilseed Markets*. Palgrave Studies in Agricultural Economics and Food Policy. Palgrave Macmillan, New York, 2015.
- [23] Adrian Fernandez-Perez, Barn Frijns, and Alireza Tourani-Rad. Contemporaneous interactions among fuel, biofuel and agricultural commodities. *Energy Economics*, 58:1–10, August 2016.
- [24] Gal Hochman and David Zilberman. Corn ethanol and us biofuel policy ten years later: A systematic review and meta-analysis. 2016 Annual Meeting, July 31-August 2, 2016, Boston, Massachusetts 235467, Agricultural and Applied Economics Association, 2016.
- [25] T.-H. Chang and H.-M. Su. The substitutive effect of biofuels on fossil fuels in the lower and higher crude oil price periods. *Energy*, 35:2807–2813, 2010.
- [26] K. Cha and J. Bae. Dynamic impacts of high oil prices on the bioethanol and feedstock markets. *Energy Policy*, 39:753–760, 2011.
- [27] S. Nazlioglu. World oil and agricultural commodity prices: Evidence from nonlinear causality. *Energy Policy*, 39:2935–2943, 2011.
- [28] G.R. Timilsina, S. Mevel, and A. Shrestha. Oil price, biofuels and food supply. *Energy Policy*, 39:8098–8105, 2011.
- [29] S. Nazlioglu, C. Erdem, and U. Soytas. Volatility spillover between oil and agricultural commodity markets. *Energy Economics*, 36:658–665, 2013.
- [30] Brian Wright. Global biofuels: Key to the puzzle of grain market behavior. *Journal of Economic Perspectives*, 28(1):73–97, Winter 2014.
- [31] Colin A. Carter, Gordon C. Rausser, and Aaron Smith. Commodity storage and the market effects of biofuel policies. *American Journal of Agricultural Economics*, Forthcoming.
- [32] Miroslava Rajcaniova, Dusan Drabik, and Pavel Ciaian. How policies affect international biofuel price linkages. *Energy Policy*, 59(C):857–865, August 2013.

- [33] Dusan Drabik, Pavel Ciaian, and Jan Pokrivcak. The effect of ethanol policies on the vertical price transmission in corn and food markets. *Energy Economics*, 55:189–199, March 2016.
- [34] Dusan Drabik, Harry de Gorter, David R. Just, and Govinda R. Timilsina. The economics of Brazil’s ethanol-sugar markets, mandates, and tax exemptions. *American Journal of Agricultural Economics*, 97(5):1433–1450, October 2015.
- [35] Dusan Drabik, Harry de Gorter, and Govinda R. Timilsina. The effect of biodiesel policies on world biodiesel and oilseed prices. *Energy Economics*, 44:80–88, July 2014.
- [36] Harry de Gorter, Dusan Drabik, and David R. Just. How biofuels policies affect the level of grains and oilseed prices: Theory, models, and evidence. *Global Food Security*, 2(2):82–88, July 2013.
- [37] R. Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B - Condensed Matter and Complex Systems*, 11(1):193–197, September 1999.
- [38] Joseph B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50, February 1956.
- [39] 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.
- [40] W.-M. Song, T. Di Matteo, and T. Aste. Hierarchical information clustering by means of topologically embedded graphs. *PLoS ONE*, 7(3):e31929, 2012.
- [41] R. Gramatica, T. Di Matteo, S. Giorgetti, M. Barbiani, D. Bevec, and T. Aste. Graph theory enables drug repurposing - how a mathematical model can drive the discovery of hidden mechanisms of action. *PLoS ONE*, 9(1):e84912, 2014.
- [42] N. Musmeci, T. Aste, and T. Di Matteo. Risk diversification: a study of persistence with a filtered correlation-network approach. *Journal of Network Theory in Finance*, 1(1):1–22, 2015.
- [43] M. Tumminello, F. Lillo, and R.N. Mantegna. Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior and Organization*, 75(1):40–58, 2010.
- [44] Ingrid Daubechies. *Ten Lectures on Wavelets*. SIAM, 2004.
- [45] L. Aguiar-Conraria, L. Azevedo, and M. Soares. Using wavelets to decompose the time-frequency effects of monetary policy. *Physica A*, 387:2863–2878, 2008.
- [46] A. Rua. Measuring comovement in the time-frequency space. *Journal of Macroeconomics*, 32:685–691, 2010.

- [47] Lukas Vacha, Karel Janda, Ladislav Kristoufek, and David Zilberman. Time-frequency dynamics of biofuel–fuel–food system. *Energy Economics*, 40(C):233–241, 2013.
- [48] Lukas Vacha and Jozef Barunik. Co-movement of energy commodities revisited: Evidence from wavelet coherence analysis. *Energy Economics*, 34(1):241–247, January 2012.
- [49] Christopher Torrence and Gilbert P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79:61–78, 1998.
- [50] C. Torrence and P.J. Webster. The annual cycle of persistence in the El Niño–Southern Oscillation. *Quarterly Journal of the Royal Meteorological Society*, 124(550):1985–2004, 1998.
- [51] Aslak Grinsted, J C Moore, and S Jevrejeva. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, 11:561–566, 2004.
- [52] Ladislav Kristoufek, Karel Janda, and David Zilberman. Co-movements of ethanol related prices: Evidence from Brazil and the USA. *GCB Bioenergy*, 8(2):346–356, March 2016.
- [53] Hrvoje Mihanovic, Mirko Orlic, and Zoran Pasric. Diurnal thermocline oscillations driven by tidal flow around an island in the Middle Adriatic. *Journal of Marine Systems*, 78, Supplement(0):S157 – S168, 2009. Coastal Processes: Challenges for Monitoring and Prediction.