

# Correlation clustering: analysis of major agricultural commodity markets

J. SZENDERÁK

University of Debrecen, Faculty of Economics and Business, Department of Agricultural Economics,  
szenderak.janos@econ.unideb.hu

*Abstract. The aim of this article is to compare the clusters formed by the correlation distances between the agricultural and the energy commodity price returns in different periods of time. The energy and agricultural markets have become more interlinked in the past ten years, which can be attributed partly to the increased usage of biofuels. According to the results of this research, after the global financial and economic crisis of 2008/09, the relationship has become tighter between the agricultural commodity prices and the price of the crude oil. Based on the hierarchical clustering, the relationship between crude oil and sugar, and especially between crude oil and vegetable oils has become stronger. These results support the hypothesis of a more interconnected agricultural and energy market after 2013. Furthermore, the emerged relationship of crude oil with the vegetable oils may indicate the connecting role of biofuels, since biofuels require agricultural input materials, partly vegetable oils. However, the role of biofuels in the present analysis requires further researches.*

## Introduction

The aim of this research is to compare the clusters formed by the correlation distances between agricultural and energy commodity price returns in different periods of time. It is widely accepted nowadays that the linkages between the energy and the agricultural commodity market have been tightened after 2008. In order to analyze the relationship between the agricultural and energy commodity prices, the correlation distances between the price returns were calculated and presented via dendrograms. The dramatic price surge in 2008/09 and again between 2011 and 2013 raised serious concerns about the behavior of the commodity prices. The primary commodities are staple foods for human consumption but they serve as animal feed as well. Thus, high food prices can have a quick and adverse effect on the households and the stakeholders of the agricultural value chain. The price changes can be either change in the level of the prices or changes in the volatility of the prices. More interlinked markets mean higher volatility transmission as well, thus the market risk and uncertainty can increase over time. This study contributes to the field by analyzing the food and energy market linkages in different subperiods between 2003 and 2018, including the most recent period as well. The idea and the methodology are motivated by Kristoufek et al. [1] and Mantegna [2] with an extended selection of commodities and sample period. With an increased focus on the relationship between the food and the crude oil prices, the results of this study offer further evidence on the more interlinking behavior of the food and the energy commodity market.

## 1. Methods

The basis for the analysis is the calculated correlation matrices, which requires stationary time series. Weak stationarity means that the expected value and the variance of the time series are independent of time, while autocorrelation depends solely on the distance between the time periods. To test stationarity, the Augmented Dickey-Fuller test (ADF) of Dickey and Fuller [3] was used according to the testing strategy of Pfaff [4]:

$$\Delta y_t = \beta_1 + \beta_2 t + \pi y_{t-1} + \sum_{j=1}^k \gamma_j \Delta y_{t-j} + u_t \quad (1)$$

Where  $y_t$  is the time series at time  $t$ , the first difference is  $\Delta y_t = y_t - y_{t-1}$ ,  $t$  is the linear trend,  $u_t$  is an error term and  $\beta_1, \beta_2, \pi, \gamma_j$  are parameters to be estimated. The parameter  $\pi$  can be defined as  $\pi = \theta - 1$ , which is more suitable in this case (see Pfaff [4]). The  $\theta$  is the parameter of the first lag of the time series. If the parameter is not significantly different from 1 (*thus*  $\pi = 0$ ), then the time series contains a unit root and it is non-stationary. Thus, our primary interest is to test whether  $H_0: \pi = 0$  or  $H_1: \pi < 0$ , thus  $\theta < 1$ .

Most of the time, the logarithmic price returns ( $\log(\Delta y_t) = \log(y_t) - \log(y_{t-1})$ ) are stationary time series. Thus, the level of the series are tested and in the case of non-rejection of the null hypothesis, the first differences are tested<sup>1</sup>. In the next step, the Correlation matrices denoted by  $M_C$  were calculated. The coefficients in the matrices are the Pearson correlation coefficients. The correlation coefficients are similarity measurements. In order to see the dissimilarity or distance between the variables, the Correlation matrices were transformed into Distance matrices  $M_D$  with the formula (*equation 3*) of Mantegna [2]:

$$M_C = \begin{bmatrix} \hat{\rho}_{11} & \cdots & \hat{\rho}_{1j} \\ \vdots & \ddots & \vdots \\ \hat{\rho}_{i1} & \cdots & \hat{\rho}_{ij} \end{bmatrix} \rightarrow M_D = \begin{bmatrix} d(1,1) & \cdots & d(1,j) \\ \vdots & \ddots & \vdots \\ d(i,1) & \cdots & d(i,j) \end{bmatrix} \quad (2)$$

$$d(i,j) = \sqrt{2(1 - \hat{\rho}_{ij})} \quad (3)$$

Where  $\hat{\rho}_{ij}$  is the Pearson correlation coefficient between the commodity  $i$  and  $j$  while  $d(i,j)$  denotes the distance between the pairs of commodities  $i$  and  $j$ . This distance formula can be seen as a metric, since it satisfies the three axioms of a metric distance (thus,  $d(i,j) = 0$ , if  $i = j$ ,  $d(i,j) = d(j,i)$  and  $d(i,j) \leq d(i,k) + d(k,j)$ ). The metric attains its minimum at zero (strong positive correlation) and its maximum at 2 (*strong negative correlation*), while in the case of no correlation between the variables, its value is  $\sqrt{2}$ . Finally, the Distance Matrices,  $M_D$  were represented on dendrograms in order to visualize the clusters formed by the commodity prices. The data was obtained from the so called Pink

<sup>1</sup> The detailed testing strategy can be found in [4], only the main results are reported here.

Sheet of the World Bank [5]. The commodity prices are nominal, monthly prices (*crude oil average, US and Europe natural gas, palm oil, soybeans, soybean oil, soybean meal, rapeseed oil, sunflower oil, maize, sorghum, US SRW and HRW wheat, beef, chicken meat and world sugar prices*) from January 2003 to January 2018 ( $T=181$  months).

The data was divided into three subcategories in order to compare different time periods. The subcategories cover the period before, during and after the food crisis, where every category contain approximately 60 months of data: the first period is from January 2007 to December 2007, the second period is from January 2008 to December 2012 and the third period is from January 2013 to January 2018. All of the calculations were performed in R (*with the packages `urca` and `corrplot`*).

## 2. Literature review

Since 2008/09, there has been a tighter link between energy and agricultural commodity markets, where price co-movements have been stronger and price volatility has been higher. The strengthened relationship was attributed mostly to the increased demand for food products as an input for biofuels.

Biofuels have gained attention after the oil crisis in the 1970s. The reason for the increasing demand nowadays is twofold: there is a need to reduce the dependency on fossil fuels and to increase the share of renewable alternatives [6]. Increasing demand for biofuels affects directly the food chain through the higher share of food crops as feedstock. Furthermore, it alters the reallocation of production resources in some way (land and water) [7]. Commodity prices are correlated with crude oil prices due to the production and transportation cost, furthermore, crop usage as biofuel input may have a role as well. Crop production is more closely linked to oil and energy prices than livestock prices [8]. Gardebroek and Hernandez [9] remark that the increased demand of bioethanol raised concerns about the stronger relationship between agricultural commodity and energy prices, which was reinforced by the market events of 2008/09 and 2010/11. This fact is due to the agricultural-based production of bioethanol and the possible effects of the higher fuel prices on agricultural commodity prices. Biofuel policy may increase food price volatility through crude oil prices. Increased demand for food commodities as biofuel inputs may tighten the linkage between crude oil and agricultural commodity prices, thus, may increase agricultural price volatility. However, biofuel production is rather driven by government mandate requirements and not by profit consideration, thus not sensitive to oil price changes [10]. Furthermore, Gilbert and Mugeru [11] found that increased volatility of grains in 2008/09 was partly due to the increased transmission of crude oil shocks, but the effect of the increased transmission was modest. The increased pass-through was accounted for 2-3% of the 10-15% volatility increase in the analyzed period (however, in the last quarter of 2008, it was more than 10%). Biofuel production partly explains the higher food price volatility in the recent decade. A comprehensive review about the role of the biofuels its relation to the land use can be found in Popp et al. [12] and in Oláh et al. [13].

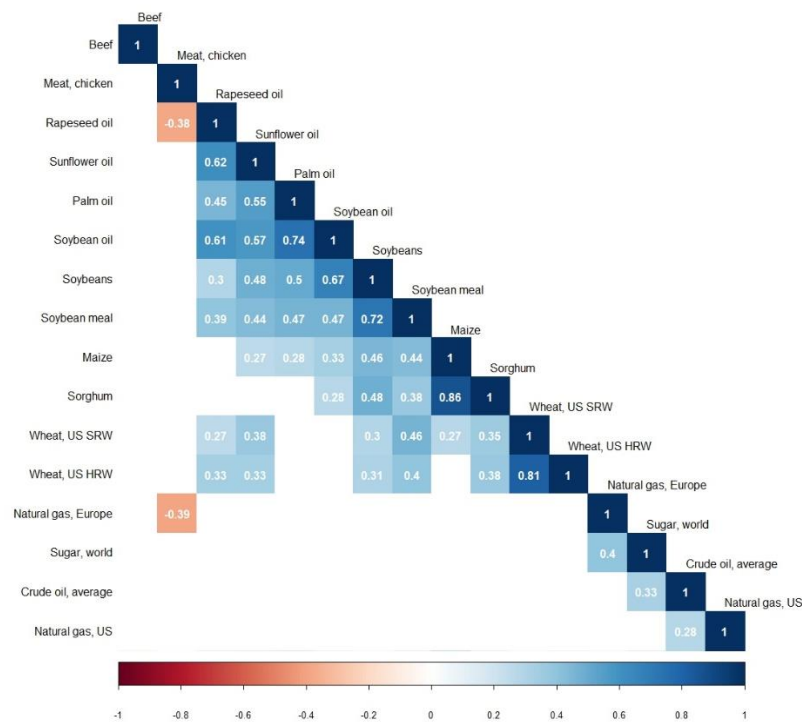
There are several types of research which provide evidence about the relationship of these markets. Kristoufek et al. [1] analyzed the relationship between the energy and the food-related commodities in the pre-crisis (2003 – 2007) and in the post-crisis period (2007 – 2011) and found tighter connections

in the second period. Ethanol has become more interrelated to corn, while biodiesel remained far from the system. Thus, the relationship may depend on the type of biofuel. Tabak et al. [14] remark that until 2008, the different types of commodities formed distinct, homogenous clusters, namely energy, agricultural and metal clusters. Furthermore, based on different measurements, Tabak et al. [14] concluded, that agricultural commodities had the most important role in the network formed by commodities. Du et al. [15] found that after 2006, there was a stronger linkage between crude oil and agricultural commodity market, and crude oil price shocks triggered the sharp price changes on the wheat and corn market. The higher amount of corn used as an input for biofuel may have contributed to the rising agricultural prices.

Kristoufek et al. [16] have found that the interaction was low between the energy and agricultural market on the short-term, but ethanol was more connected to the food clusters on the medium-term, while biodiesel remained in the fuel branch. When the oil prices were high, biodiesel prices were driven by other fuel prices, while ethanol prices were driven mostly by soybeans and sugar. In the case of low oil prices, biofuels tended to remain relatively far from the system. This effect can be due to the fact, that fodder prices become more sensitive to oil prices when those are at a high level. Furthermore, Kristoufek et al. [16] have found Granger causality running from corn and wheat prices to ethanol prices and from German diesel fuel and crude oil prices to biodiesel prices. The results of Lucotte [17] support the hypothesis that crude oil and food markets are more interlinked in the post-boom period, that is after 2008/09, which can be explained by the oil price surges and its effect on the substitution effect between biofuels and fossil fuels. Tadesse et al. [7] argue that food and energy market became more connected due to the stronger relationship between oil prices and food price spikes and volatility. López Cabrera and Schulz [6] remarks that the correlation between the crude oil and the agricultural commodity market has strengthened after 2005, which can be due to the biofuel boom. There was no visible evidence for trending volatility in the prices, but in unstable periods, volatility has increased considerably strongly. Allen et al. [18] found evidence also for the relationship between food commodities and energy markets, although the market volatility conditions affected the strength and the direction of these relationships. Impulse response changes to shocks were dependent on the low or high phases of volatility. Saghaian et al. [19] showed that lagged corn and ethanol price volatility has no effect on the current oil price volatility. However, there were volatility spillovers from oil prices to corn and ethanol prices and from ethanol prices to corn prices. Furthermore, there was evidence for bidirectional volatility spillovers between corn and ethanol. Corn market responses were dependent on the direction of the price changes, thus there was asymmetric volatility transmission between food and biofuel market.

Until 2009, the majority of the commodity prices have clearly increased. Between 2007/08 and 2010/11, the global financial crisis and the adverse weather-related events resulted in a great turmoil on the market with sudden price changes. Around 2011, commodity prices have normalized, but volatility higher volatility compared to the periods before the crisis. Furthermore, the trend of the prices was rather decreasing, while some of the commodities possessed remarkably similar movements. In general, volatility of the commodity prices has increased after the crisis. While the correlations between them became more significant, their movements became tighter (*Graph 1, 2, 3 and 4 in the Appendix*). In the first period, a given energy or food commodity was correlated mostly

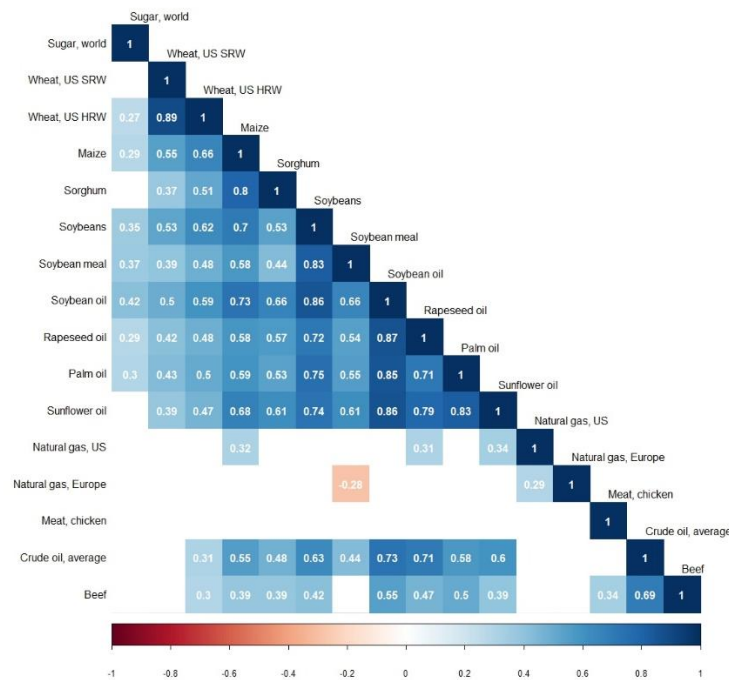
with other energy or food commodities, respectively. There was a relatively strong and significant correlation between the different types of oilseeds and vegetable oils. The correlation coefficient was negative between the rapeseed oil and the chicken meat prices, furthermore between the natural gas and the chicken meat prices. Sugar prices exhibited a moderately strong relationship with the EU natural gas prices, while beef prices were uncorrelated with any of the other commodities in the system. Thus, the linkages between the food and energy market were not visible in the first period (Graph 1.).



Graph 1.: The correlation coefficients in the first period

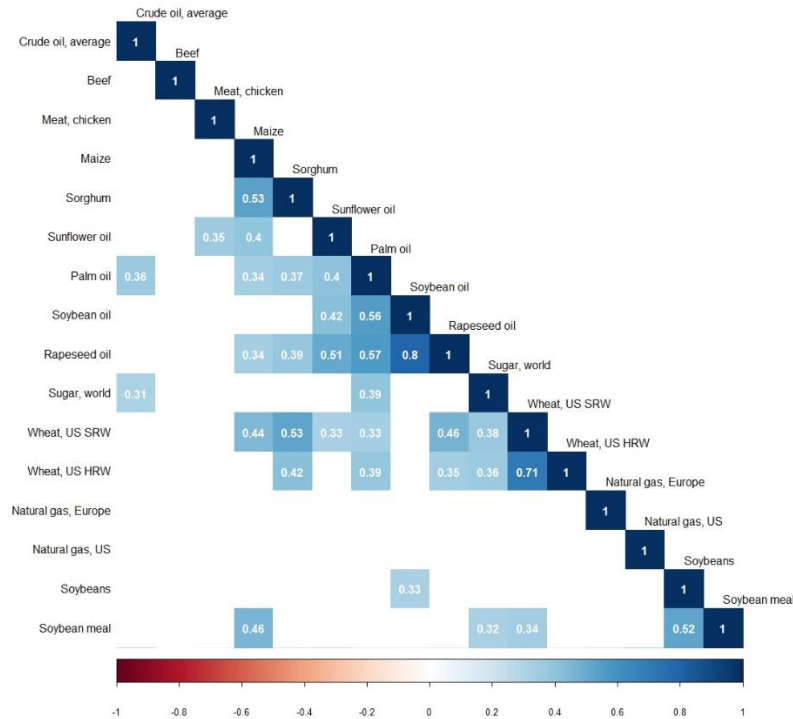
Source: Own calculation (2018)

In the second period, the global financial and economic crisis and the adverse weather conditions in some part of the world have caused sudden price changes on the market. The sudden rise and fall in the commodity prices were reflected in the higher correlations between the variables. This fact was observed by Huchet-Bourdon [8] as well. The relationships were very unusual compared to the first period. Most importantly, crude oil prices were correlated with the majority of the commodity prices being analyzed. Beef prices were correlated with the cereal, oilseeds and vegetable oil, crude oil and chicken meat prices, while sugar and US SRW wheat prices did not exhibit any significant relationship with the rest of the system. The only significant negative correlation was between the EU natural gas and soybean meal prices (Graph 2.).



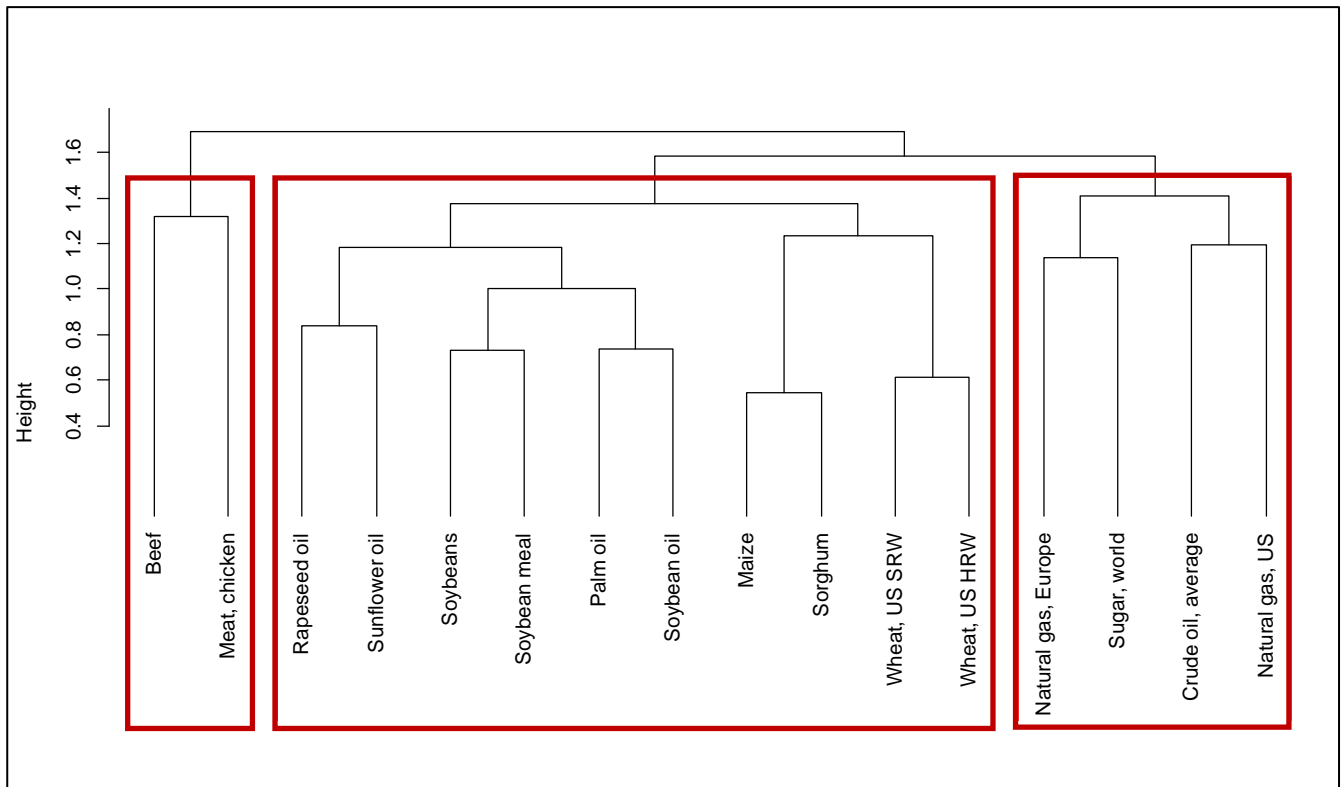
Graph 2.: The correlation coefficients in the second period  
 Source: Own calculation (2018)

The third part was the period after the global financial crisis and the following price normalization. After the price normalization, the nature of the relationships was weakened or disappeared. However, significant relationship emerged between the food and the agricultural commodity market. Palm oil and sugar prices showed moderate correlation with the crude oil prices. Although the linkages between the natural gases and the food commodities have disappeared, a significant correlation has developed between the crude oil and the palm oil prices. Although the results are reported at the 5 percent level of significance, more such a connection can be detected at the 10 percent level of significance. This fact implies that the relationship between the food and the energy market has tightened (Graph 3.).



Graph 3.: The correlation coefficients in the third period  
 Source: Own calculation (2018)

The correlation coefficients can be an indicator of the underlying market behavior, while hierarchical clustering may offer a more comprehensive picture. The disadvantage of the method that the chosen number of clusters highlighted is arbitrary. Three clusters were highlighted mostly due to their separated branches. The aim of the hierarchical clustering was to see whether the energy and the food prices appear in the same cluster. There is a clear distinction between the groups. Three types of branches were formed in the first period. The first one is a meat branch, which included the chicken and the beef meat. The second branch was formed by the energy commodity prices and the sugar world prices. This kind of distinction was clearly indicated by the historical price movements as well. The meat and the sugar world prices had their own unique movements all the way along. The biggest cluster is the pure food cluster, which is separated into a cereal branch and an oilseed and vegetable oil branch. Thus, even in the first period, there was only one commodity (*sugar*), which had a strong relationship with energy prices, even stronger than with the rest of the food prices (*Graph 4.*).

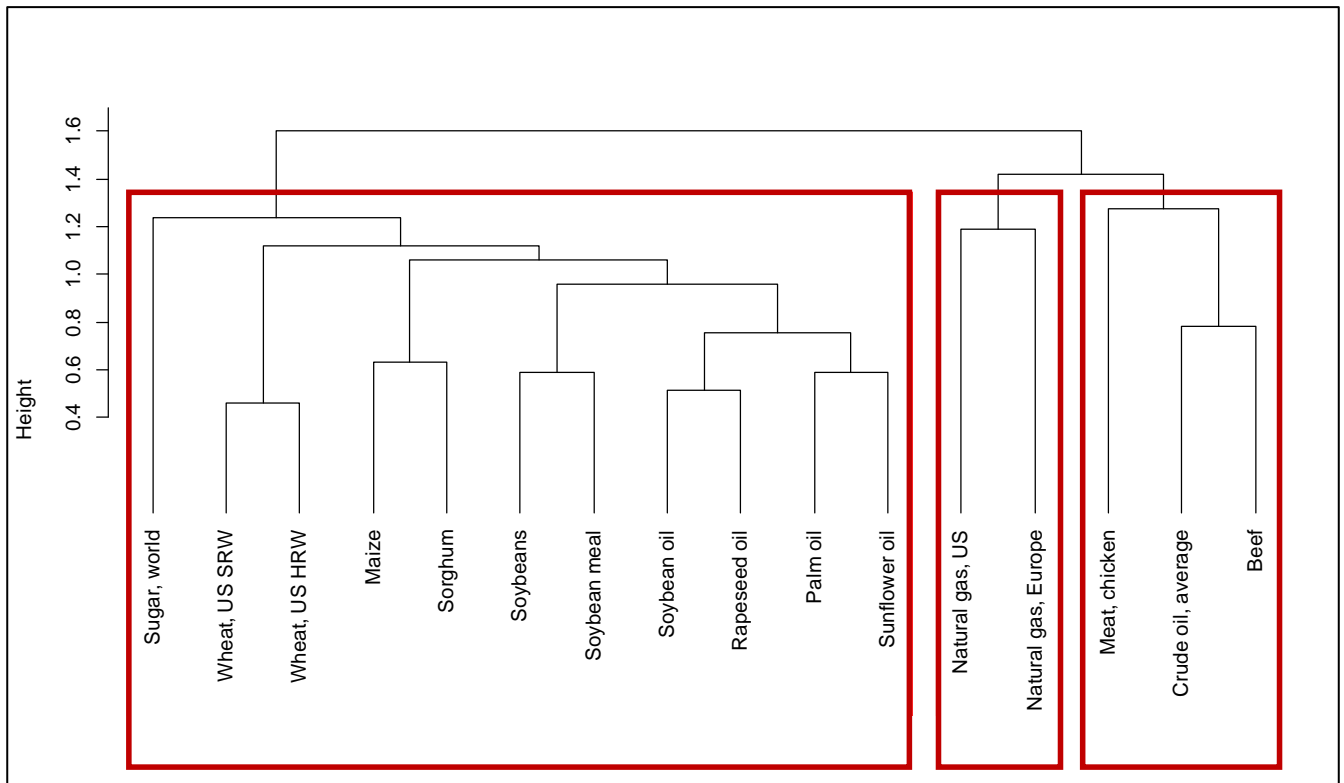


Graph 4.: The commodity cluster in the first period

Source: Own calculation (2018)

In the second period, the system has clearly changed considerably compared to the period before. Sugar was included in the food branch, but still, it was relatively far from the rest of the system. The rest of the food branch did not show any extreme connection. Cereal prices, oilseed and vegetable oils formed their own separate branch, but overall, the commodities in the food cluster were strongly connected. The meat cluster included the crude oil prices in this period, which indicates some kind of connection between the food and the energy market, even without the sugar world prices. The US and EU natural gas formed their own branch. Interestingly, the correlation coefficients indicated that crude oil became much more correlated with the rest of the system for the second period. Intuitively, we would expect the crude oil prices to be much more connected to the rest of the system due to the stronger correlation. While this was true, at the same time food commodities have developed much stronger connections as well, especially those which are closely related by nature. Thus, hierarchical clustering formed still the separate branches (*Graph 5*).

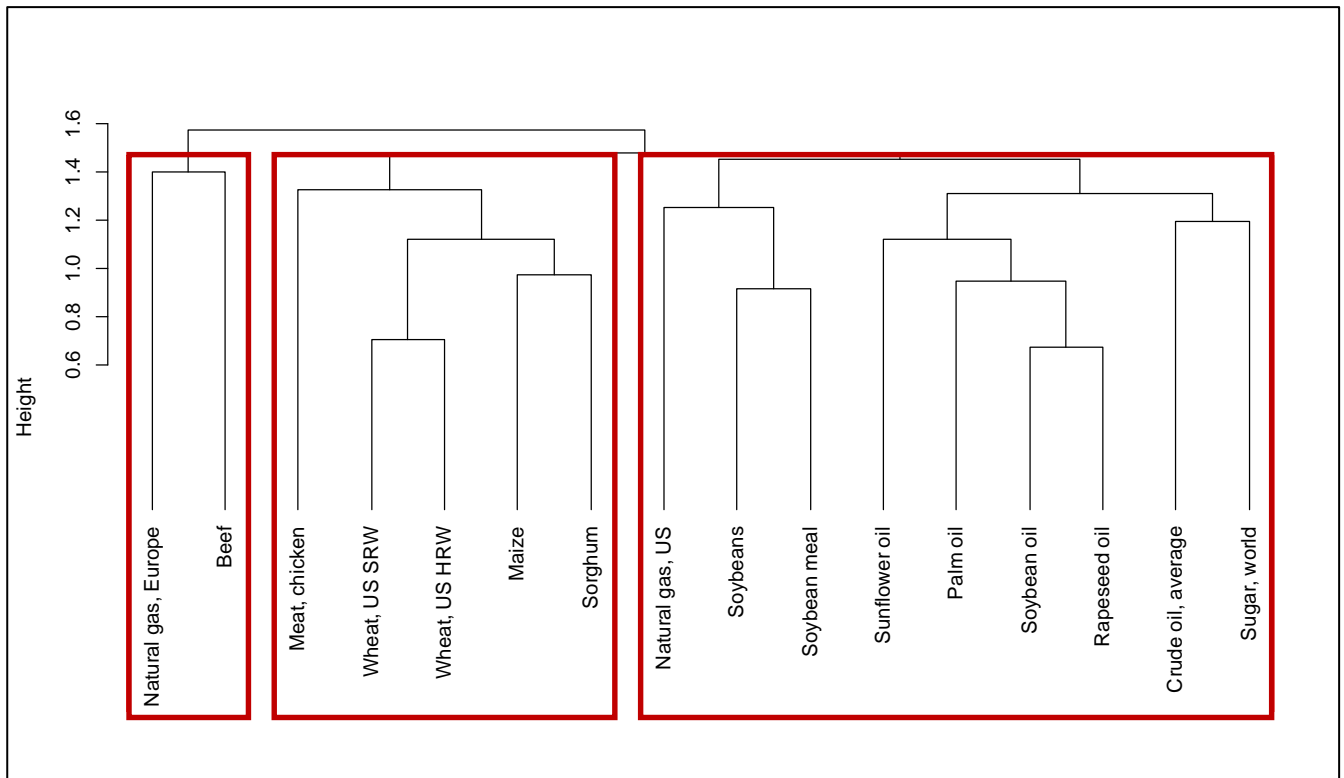




Graph 5.: The commodity cluster in the second period

Source: Own calculation (2018)

Finally, in the last period, the main interest is whether the food and agricultural market have some connection which did not exist before. The food branch included just a few commodities in the third period. The US SRW and HRW wheat, the maize, the sorghum and surprisingly, the chicken meat formed this branch. There was also a small branch which included the EU natural gas and the beef meat. This latter branch indicates that there is a connection between the agricultural and energy market in the recent years. However, beef prices have behaved very differently compared to the rest of the commodities, thus it may be not a good indicator of the emerging relationship. The oilseeds and the vegetable oils were not included in a common branch with energy commodities before 2013. After 2013, however, crude oil prices formed a tighter linkage with these agricultural commodities, especially with the vegetable oil prices. Furthermore, the significant relationship between crude oil prices and world sugar prices has still remained. These results support the hypothesis, that the agricultural and the energy market are more interconnected in the recent years. Furthermore, the tight linkage with the vegetable oil prices may indicate the connecting role of biofuels. However, this latter argument requires further researches (Graph 6.).



Graph 6.: The commodity cluster in the third period

Source: Own calculation (2018)

## Conclusion

The energy market can affect the agricultural market in several ways, especially through high and more volatile crude oil prices. After the oil price boom in the 1970s, the recent price boom around 2008/09 and between 2011 – 2013 have gained a considerable attention again. The emerging crisis has coincided with the development of the biofuels as an alternative to the fossil fuels. Some researchers argued that the biofuel-related policies contributed to the price increases on the agricultural market. The related results are highly mixed and dependent on several factors. Most of the researchers have found that there is a tighter linkage between the energy and the agricultural commodity market. The role of biofuels in the price increases have remained unclear together with other factors, like the role of financialization of the agricultural market and the speculation. The results of this research support the hypothesis of the more interrelated energy and agricultural market. The connections have become tighter, especially between the crude oil and the vegetable oil prices. The emerged relationship may indicate the possible connecting role of biofuels, however, a further argument cannot be based on this research.

## Acknowledgments

I would like to thank the reviewer for the comments made. The research was supported by EFOP3.6.3-VEKOP-16-2017-00007- "Young researcher for talent" –Supporting career in research activities in the higher education project. The author furthermore receives financial support from the Pallas Athéné Domus Educationis (PADE) foundation.

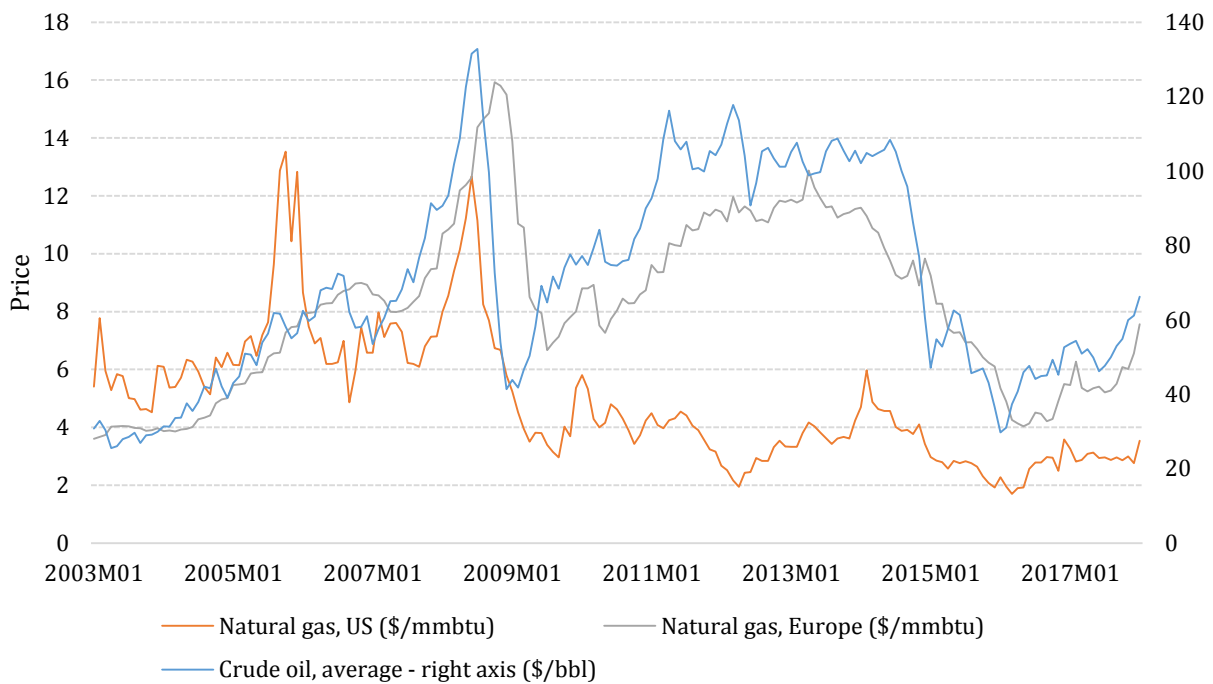
## References

- [1] L. Kristoufek – K. Janda – D. Zilberman (2012) *Correlations between biofuels and related commodities before and during the food crisis: A taxonomy perspective*. Energy Economics, 2012. 34 (5) pp. 1380-1391.
- [2] R. N. Mantegna (1999) *Hierarchical structure in financial markets*. The European Physical Journal B-Condensed Matter and Complex Systems, 11 (1) pp. 193-197.
- [3] D. A. Dickey – W. A. Fuller (1981) *Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root*. Econometrica, 49 (4) pp. 1057-1072.
- [4] B. Pfaff (2008) *Analysis of integrated and cointegrated time series with R*. Springer Science & Business Media.
- [5] World Bank (2017) *The World Bank Commodity Markets Pink Sheet*. The World Bank Group.
- [6] B. López Cabrera – F. Schulz (2016) *Volatility linkages between energy and agricultural commodity prices*. Energy Economics, 2016. 54 pp. 190-203.
- [7] G. Tadesse, et al. (2014) *Drivers and triggers of international food price spikes and volatility*. Food Policy, 47 pp. 117-128.
- [8] M. Huchet-Bourdon (2011) *Agricultural Commodity Price Volatility: An Overview*. In: OECD Food, Agriculture and Fisheries Working Papers. OECD. p. 48.
- [9] C. Gardebroek – M. A. Hernandez (2013) *Do energy prices stimulate food price volatility? Examining volatility transmission between US oil, ethanol and corn markets*. Energy Economics, 40 pp. 119-129.
- [10] C. L. Gilbert – C. W. Morgan (2010) *Food price volatility*. Philos Trans R Soc Lond B Biol Sci, 2010. 365 (1554) pp. 3023-34.
- [11] C. L. Gilbert – H.K. Mugeru (2014) *Food Commodity Prices Volatility: The Role of Biofuels*. Natural Resources, 5 (5) pp. 200-212.
- [12] J. Popp, et al. (2014) *The effect of bioenergy expansion: Food, energy, and environment*. Renewable and Sustainable Energy Reviews, 32 pp. 559-578.
- [13] J. Oláh, et al. (2017) *The Role of Biofuels in Food Commodity Prices Volatility and Land Use*. Journal of Competitiveness, 9 (4) pp. 81-93.
- [14] B. M. Tabak – T. R. Serra – D .O. Cajueiro (2010) *Topological properties of commodities networks*. The European Physical Journal B, 74 (2) pp. 243-249.
- [15] X. Du – C. L. Yu – D. J. Hayes (2011) *Speculation and volatility spillover in the crude oil and agricultural commodity markets: A Bayesian analysis*. Energy Economics, 33 (3) pp. 497-503.
- [16] L. Kristoufek – K. Janda – D. Zilberman (2013) *Regime-dependent topological properties of biofuels networks*. The European Physical Journal B, 86 (2) p. 40.

- [17] Y. Lucotte (2016) *Co-movements between crude oil and food prices: A post-commodity boom perspective*. Economics Letters, 147 pp. 142-147.
- [18] D. E. Allen, et al. (2018) *A cointegration analysis of agricultural, energy and bio-fuel spot, and futures prices*. Applied Economics, 50 (7) pp. 804-823.
- [19] S. Saghaian, et al. (2018) *Asymmetric Price Volatility Transmission between US Biofuel, Corn, and Oil Markets*. Journal of Agricultural & Resource Economics, 43 (1).

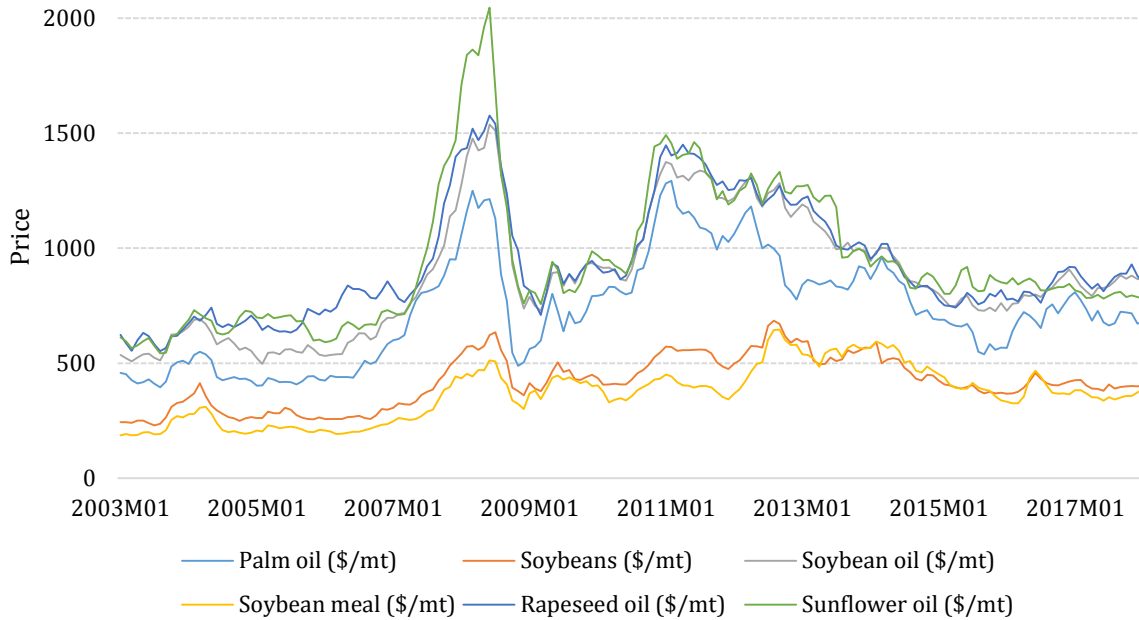
## Appendix

### Appendix 1.



*Graph 1.: The development of energy prices  
Sources: Based on the data of the World Bank [5]*

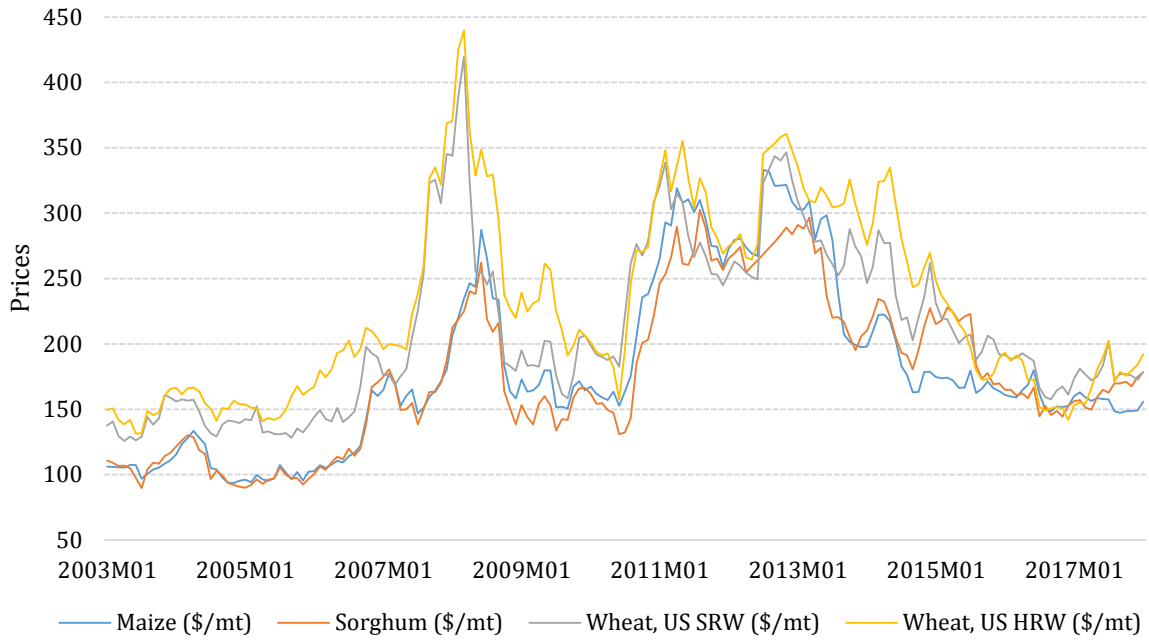
## Appendix 2.



Graph 2.: The development of oilseeds and vegetable oil prices

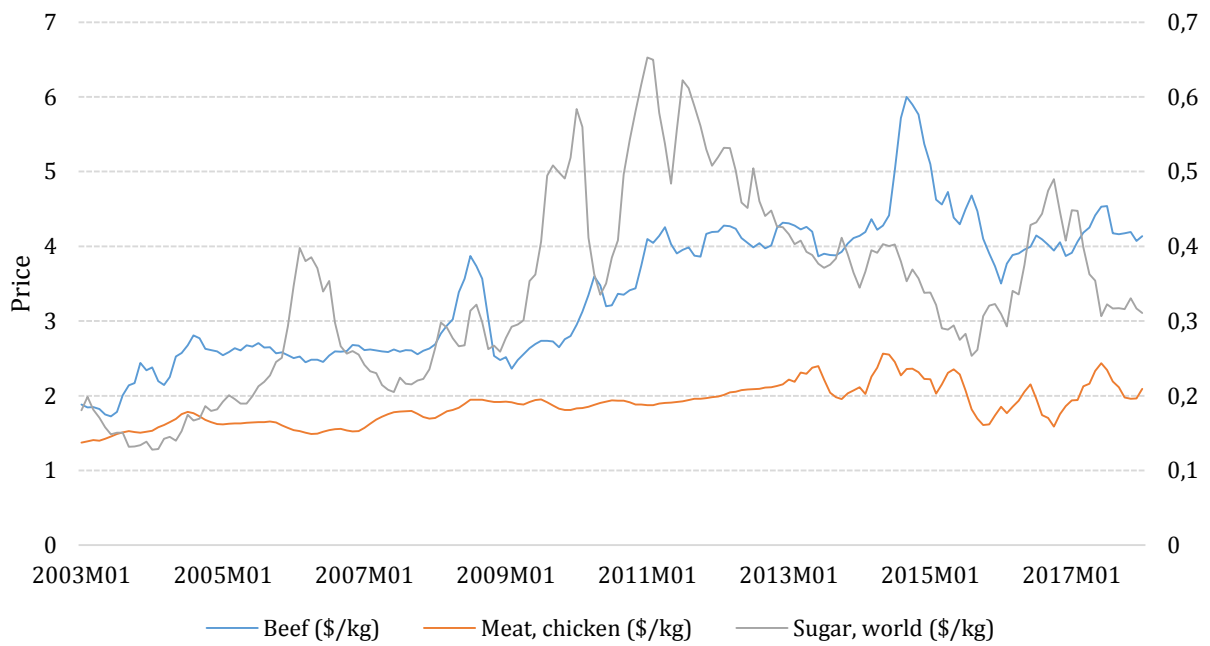
Sources: Based on the data of the World Bank [5]

### Appendix 3.



*Graph 3.: The development of cereal prices*  
*Sources: Based on the data of the World Bank [5]*

## Appendix 4.



Graph 4.: The development of beef meat and sugar prices

Sources: Based on the data of the World Bank [5]