CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Impact of Drought on Commodity Market Forecasting

Bachelor thesis

Author: Kristian Racocha
Study program: Economics and Finance
Supervisor: Mgr. Marek Hauzr
Year of defense: 2020

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, May 7, 2020

Kristian Racocha

Abstract

This thesis analyzes the relationship of drought (SPEI-01) and commodity prices. Cyclical properties of both components are extracted and by utilizing the spectral factor models proposed by Bandi et al. (2019) the impact of SPEI-01 on various commodity prices across a range of frequencies is examined. The analysis reveals that most of the commodities are impacted by SPEI-01 when considering cycles of lengths $2-4$ months and $32-64$ months. For frequencies, where a significant relationship is found, the commodity prices are forecasted using the ARIMA model with and without the inclusion of the drought variable. The results suggest that the inclusion of drought is beneficial for a majority of the models, however, in a considerable number of occasions the models exhibited a diminished performance.

Abstrakt

Tato práce zkoumá vztah mezi suchem (SPEI-01) a cenami komodit. Pomocí spektrálních faktorových modelů, které navrhli Bandi et al. (2019), je možné analyzovat cyklické vlastnosti obou složek a následně pozorovat, jaký dopad má SPEI-01 na ceny komodit v různých frekvencích. Tento rozbor poukazuje na kauzální vztah mezi proměnnými pro většinu komodit, konkrétně v cyklech v rozmezí $2-4$ měsíců a $32-64$ měsíců. Ve frekvencích, kde byl odhalen signifikantní vztah mezi proměnnými, jsou ceny komodit předpovězeny pomocí dvou ARIMA modelů, jedním, který do modelu zahrnuje sucho, a druhým, který využívá pouze časovou řadu cen. Výsledky ukazují, že ve většině případů zahrnutí sucha přispělo k lepší předpovědi, ovšem pro nezanedbatelné množství případů se naopak výkonnost modelu zhoršila.

JEL Classification	Q02, Q54
Keywords	spectral factor models, commodity market, drought, forecasting

Title	Impact of Drought on Commodity Market Forecasting

Author's e-mail	`racochak@gmail.com`
Supervisor's e-mail	`marek.hauzr@gmail.com`

Acknowledgments

I would like to thank to Mgr. Marek Hauzr for his guidance and patience and to my family and friends for their long lasting support.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record

Racocha, Kristian: *Impact of Drought on Commodity Market Forecasting.* Bachelor thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2020, pages 57. Advisor: Mgr. Marek Hauzr

Contents

List of Tables

List of Figures

List of Acronyms

ADF Augmented Dickey-Fuller test

AIC Akaike Information Criterion

AR autoregressive

ARIMA autoregressive integrated moving average

ARIMAX autoregressive integrated moving average exogenous

CLT Central Limit Theorem

DHT discrete Haar transform

FEMA Federal Emergency Management Agency

f.o.b. Free On Board

HRW hard red winter

HWT Haar Wavelet Transform

ICCO International Cocoa Organization

IPCC Intergovernmental Panel on Climate Change

MA moving average

MAE mean average error

MAPE mean average percentage error

mt metric ton

NCEI National Centers for Environmental Information

OLS Ordinary Least Squares

PDSI Palmer Drought Severity Index

PET potential evapotranspiration

RCPs Representative Concentration Pathways

RMSE root mean square error

sc-PDSI self-calibrated Palmer Drought Severity Index

SPEI Standardised Precipitation Evaporation Index

SPI Standardised Precipitation Index

SR success ratio

Bachelor's Thesis Proposal

Author	Kristian Racocha
Supervisor	Mgr. Marek Hauzr
Proposed topic	Impact of Drought on Commodity Market Forecasting

Author's name and surname : Krisitan Racocha

mail : racochak@gmail.com

Phone : 723479650

Supervisor's name : Mgr. Marek Hauzr

Supervisor's email : marek.hauzr@gmail.com

Proposed Topic : Impact of Drought on Commodity Market Forecasting

Preliminary scope of work:

Research question and motivation The main research question for this thesis is to find the impact that weather, specifically drought, has on the commodity cycles and the ability to forecast their future prices.

Cycles are key to understanding the economy, and commodity prices also exhibit cyclical movements. The first to notice such behavior was Clément Juglar (1862) as he identified a 7 to 11 year investment cycle, followed by a description of short 3 to 4 year cycles (Kitchin 1923), long waves with a periodicity of roughly 50 years (Kondratieff 1925), and 17 to 30 year cycles (Kuznets 1930). Scientists continue to build on the research and adapt the theory to modern data.

With knowledge about the commodity cycles, economists are able to predict future prices with greater precision. The forecasting models analyze previously observed time series data in order to predict future values. The first systematic approach to forecasting time series data was the introduction of the autoregressive integrated moving average (ARIMA) (Box and Jenkins 1970). According to a study, drought levels have been increasing for the past 200 years (Briffa, van der Schrier and Jones 2009). In addition to this, a paper by the journal Nature Climate Change (Park, Jeong, Joshi, et al. 2018) predicts major aridification in the coming years as an implication of rising temperatures.

My thesis will study both the cyclical nature of commodity prices with an oversight to the drought situation. Evidence suggests that drought levels have an impact on commodity prices in a certain time period (Trostle 2011) and I will add to this research with a long-term view of the relation between the two variables.

Contribution The purpose of this thesis is to further the current research regarding interrelation of commodity prices and global drought. The contribution will be to observe the impact that drought has on forecasted commodity prices using time series analysis and to help clarify whether including the extra variable in the model brings more accurate results.

Methodology The research will be focusing on long-term data of prices of agriculture commodities and SPEI ("Standardized Precipitation-Evapotranspiration Index"), which is a drought index combining precipitation and temperature data. Firstly, I will try to prove a dependency between the two variables, agriculture commodity price changes and the SPEI data using a comparative approach. Secondly, I will create forecasting models. A benchmark one, featuring solely the commodity price data, and a model incorporating the drought index into the calculations. In the next step I will compare the two outcomes and conclude, whether including drought as a variable leaves us with a more realistic estimate.

Outline

1. Introduction
2. Literature review
3. Data description
4. Methodology
5. Discussion over the results
6. Conclusion

List of academic literature:

Bibliography

- Briffa, K., Schrier, G., and Jones, P. (2009). Wet and dry summers in Europe since 1750: evidence of increasing drought. International Journal of Climatology, 29:1894−1905.
- Grinin, L., Devezas, T. C., and Korotayev, A. (2014). Kondratieff waves: Juglar−Kuznets−Kondratieff; yearbook.
- Park, C.-E., Jeong, S.-J., Joshi, M., Osborn, T. J., Ho, C.-H., Piao, S., Chen,D., Liu, J., Yang, H., Park, H., Kim, B.-M., and Feng, S. (2018). Keeping global warming within 1.5°C constrains emergence of aridification. Nature Climate Change, 8(1):70−74.
- Trostle, R. (2011).Why Have Food Commodity Prices Risen Again? DIANEPublishing Company.
- Tsay, R. (2000). Time series and forecasting: Brief history and future research. Journal of The American Statistical Association, 95:638–643.

Author Supervisor

Chapter 1

Introduction

Meteorological extremes have intensified, appeared with greater frequencies, and are now longer in duration in recent years according to IPCC (2012). Temperature patterns of the past 2000 years reconstructed by Neukom et al. (2019) report a coherent increase in temperature in the past 150 years (after the Industrial Revolution) and find evidence of the twentieth century being the warmest of the past two millennia across the entire globe. The increase in temperatures has become more prevalent since the 1970s as shown by Folland et al. (2006) and is attributed to a major increase in production of greenhouse gases (see, e.g., Mitchell and Karoly (2001), Solomon et al. (2010)). The report from World Meteorological Organization (2014) shows that each of the past three decades has been warmer than the last, culminating with 2001−2010 as the warmest decade on record since 1850. Not only has the temperature been increasing in the recent decades, but so has the frequency and intensity of other meteorological extremes, such as drought or hurricanes (see, e.g., Arndt and Blunden (2012), IPCC (2012)).

The definition of drought is a prolonged period of abnormally low rainfall, leading to a shortage of water, but there is no general agreement on a threshold. A period is deemed as a dry one, if a dry spell relative to the normal local condition is recorded. Many different types of droughts are observable, such as agricultural, meteorological, hydrological or socioeconomic drought (see Wilhite and Glantz (1985)). Otkin et al. (2017) characterize a rapidly developing drought with a short duration, a flash drought, which is likely to occur during growing season and cause intense crop stress. Briffa et al. (2009) reveal that drought levels have been increasing with evidence dating back to 1750 and there is now an increased risk of the duration, severity and extent of droughts as a consequence of global warming, (see Trenberth et al. (2003)). In contrast to to dry spells there is permanent aridity found in deserts.

Dai (2011) shows that since the middle of the 20th century, the extent of areas experiencing aridity and drought has increased substantially, mainly due to widespread drying since the 1970s. The areas in concern are mostly in the subtropical zone. He attributes the drying trend largely to El Niño-Southern Oscilation and changes in Atlantic sea surface temperatures, but also to the rapid warming since the 1970, which increased the atmospheric moisture demands and likely altered the atmospheric circulation patterns. He further simulates a long-term forecast of the global drought conditions and predicts a great increase of areas facing widespread drought and aridity in the upcoming decades.

The concept of Representative Concentration Pathways (RCPs) has emerged in recent years (see IPCC (2014)). They represent four possible pathways of the climate situation in 2100, ranging from RCP 2.6 as the ideal scenario, and RCP 8.5 as the worst case scenario we are heading towards if no action is taken (the current level emissions are tracking close to the RCP 8.5 scenario). Zhou and Hong (2013) simulate the projected changes to Palmer drought severity index (PDSI) under the RCP 8.5 scenario and, similarly to Dai (2011), predict that both moderate and extreme drought areas have statistically a significant increasing trend.

With the evidence of drought levels rising over the past number of decades, and the dire predictions for the future decades it appears certain to lead to deeper consequences than those already be experienced. Drought has an economic impact on ecosystems, biodiversity and water supply. It is interconnected with several industries, such as forestry, the energy sector, and agriculture (see, e.g., Kraemer (2007), Freire-González et al. (2017)).

The fact that drought has a negative impact on agriculture is generally agreed upon. According to FEMA (1995), drought causes $\$6-8$ billion per year in damages on average across the U.S. alone, however, during years with severe drought the figure is much higher (\$40 billion in 1988). A more recent report by NCEI (2020) aims to estimate the costs of weather and climate disasters and evaluate that the damages caused by drought is on average \$9.6 billion per regional event. A number of studies have examined the impact of weather conditions on agriculture commodities. Piesse and Thirtle (2009) and Baffes and Haniotis (2010) investigate the behavior of commodity prices in 2006-2008 and discuss how drought and other weather events during the period contributed to a rise in grain prices. Brooks et al. (2014) found that larger areas under extreme drought conditions lead to a larger absolute price increase. Trostle (2011) examines the sharp increase in commodity prices after the 2008 financial crisis and recognizes that in regions known for producing agriculture commodities, drought was among other factors a major catalyst in the increase of prices. Potopová et al. (2015) investigates the performance of Standardised Precipitation Evaporation Index (SPEI) at various lags for agricultural risk assessment and finds a high correlation for cereals during the May to June. Burkhead and Klink (2018) examine the relationship between crop yield of five major commodities grown in the United States and the Palmer Drought Severity Index (PDSI) and find that corn and cotton are statistically significant at a 5% confidence level.

Prices of commodities traded in commodity markets fluctuate, and just as in any other financial market, they exhibit cyclical behavior over a certain period of time. The existence of cycles is one of the key factors examined in economics and it has been studied by researchers for more than a century. The first to notice such behavior was Juglar (1862) as he identified a 7 to 11 year investment cycle, followed by a description of short 3 to 4 year cycles (see Kitchin (1923)), long waves with a periodicity of roughly 50 years (see Kondratiev (1925)), and 17 to 30 year cycles (see Kuznets (1930)). Mitchell (1927) and a group of scientists from Harvard studied cycles statistically opposed to a logical interpretation and found cycles with lengths of 3 to 4 years, to a certain extent coinciding with Kitchin cycles. Scientists continue to build on the research and adapt the theory to modern data.

Le et al. (2019) used data dating back to 1870 to find evidence of long 36-year commodity price cycles in 608 years per thousand. Labys et al. (2000) identify a predominance of two kinds of cycles in primary commodity prices, one with periodicity of less than 12 months reflecting speculative influence of future trading, and a second with a periodicity of over 24 months corresponding with basic findings on the duration of short term macroeconomic cycles in the U.S. economy.

In the book Business Cycles, Schumpeter (1939) tried to consider the the structure of medium-term cycles (Juglar cycles) to be identical with the structure of long waves (Kondratiev waves), but his attempt to establish a general theory of business cycles failed. This could be given by the fact that the business cycle was originally thought of as being deterministic in nature (see, e.g., Benhabib and Nishimura (1979; 1985), Day and Shafer (1987)). But as further research developed, economists began to understand that the cycles follow a stochastic pattern (i.e., it contains a certain level of randomness, error) or a endogenous deterministic cycle, which is subject to exogenous shocks, such as technology innovation, weather, change in consumer preferences, etc. (see, e.g., Beaudry et al. (2017), Koopman and Lee (2005)). Cashin et al. (1999) focus on identifying a consistent 'shape' to commodity cycles, but find little to no evidence supporting this idea. The show that the probability of the end of a price boom or slump is independent of the time already spent in the boom.

Therefore, the commodity price time series takes on the shape of quasi-cycles, i.e., imperfectly cyclical behavior that is accompanied with an element of uncertainty. This phenomenon is well captured by Wold's representation (1938), where a covariance-stationary time series can be written as a sum of two time series, a deterministic and a stochastic one. Following the Wold representation Bandi et al. (2019) construct their novel spectral factor models, which build on the framework of traditional factor models and decompose them for analysis in various frequencies. The extraction of the frequency-specific components allows researchers to reveal additional meaningful information. Kamara et al. (2016) investigate the pricing of risk factors across different investment horizons and emphasize their importance when determining risk premia.

However, the spectral decomposition by Bandi et al. (2019) does not exclusively apply to the pricing models of financial assets, and opens the door for an analysis across frequencies for a wide range of topics. With the changes in weather being more prominent in recent years and with commodity prices and drought both exhibiting cyclical behavior, studying the comovements of the two series may reveal valuable information. The research focuses on a regression analysis using the spectral factor models, determining at which frequencies commodity prices are most impacted by drought. This is followed by a performance comparison of forecasting models including and excluding the information about drought.

The thesis is organized as follows, in Chapter 2 the relevant literature about data transformations, frequency analysis and drought indices is reviewed, Chapter 3 follows with a description of the data. In Chapter 4 the methodology of the analysis is presented, Chapter 5 summarizes the results obtained from the models, and concluding remarks are given in the Chapter 6.

Chapter 2

Literature review

2.1 Filters and detrending

For the analysis to work, stationary data are needed for both time series used in the model (i.e., the mean and variance are constant across the entire time series). However, the raw data for commodity prices and the global drought index are not yet in the desired form because they exhibit a trend. Therefore, detrending of the data is needed to eliminate the distortions and to observe the subtrends that will reveal the cyclicality of the time series. Heavily related to the detrending of the data is also the filter applied. The same range of values that is used for the detrending is then used for the filter as well. If the two methods utilize a similar mechanism of modifying the data, they may collide and cause distortions in the series. Therefore, a review of detrending and filtering methods is needed to single out the ones appropriate for further analysis.

A time series is usually decomposed as a sum of unobservable components, a trend or growth, a cyclical, a seasonal, and an irregular (error) component (see Pedersen (2003)). The decomposition of a non-stationary time series into a trend component and a transient component, which is a stationary process with mean zero, was popularized by Beveridge and Nelson (1981). The problem arises when defining the individual components. Several differing methods on how to extract them have been developed. Canova (1998) argues that there are many different statistical representations of the trend and by choosing, which detrending method to use, researchers are de facto selecting a particular concept of a business cycle fluctuation over another. Singleton (1988) mentions that stylized facts leading to specifications of business cycle models may have been distorted by prefiltering procedures.

Wu et al. (2007) define two types of trends, predetermined and adaptive. Predetermined or extrinsic trends are viewed as a simple straight line fitted to the data, and the most common detrending process usually consists of removing the straight line best fit, yielding a zero mean residue. They argue that this method is only feasible in a purely linear and stationary world and exhibits clear limitations in real-world applications. The adaptive or intrinsic trends must be an integral part of the data and should not be affected by additional values. Yet, after detrending the data must retain its information about potential cycles. Wu et al. (2007) follow by defining a trend as "an intrinsically fitted monotonic function or a function in which there can be at most one extremum within a given data span"(p.14890). Modirroosta (2013) arrives at a similar conclusion by showing that a linear trend successfully removes a changing mean across time and although it does not alter the frequency of cycles, it struggles to remove unit roots effectively, as shown on metal prices. Much improved goodness of fit and significance is found for quadratic trends compared to the linear counterparts, while retaining the strengths of the simple linear trend.

A common way of achieving stationarity is using first order differencing (i.e., plotting changes in values instead of actual values) and consequentially second and third order differencing if needed. However, for this research, first differencing is inappropriate as it might intervene with the applied filter.

An alternative method of modifying the data is taking the relative changes in the prices. This allows to analyze the returns of the individual commodities instead of the nominal prices. It is possible that this procedure by itself will generate a covariance stationary series, but this may not be the case. Granger and Stărică (2005) explore the non-stationarity in stock market returns and relinquish the assumption of stationarity. However, after utilizing the augmented Dickey-Fuller test for stationarity, the series of commodity returns is deemed covariance stationary and thus taking relative changes functions as a detrending method. Relative changes are also less likely to intervene with the applied filter as it does not rely solely on taking the differences between the two observations.

For the drought index calculating relative changes or differences is undesirable, as the goal is not to find the impact of the changes, but rather the impact of the magnitude of drought on commodity prices. The application of relative changes to the drought index data creates a large number of outliers, because it contains many values close to zero, which only further proves the point. Therefore, a quadratic trend line will be removed from the series (if found) to achieve a constant mean over time. The variance is almost constant across the entire series, so it does not affect the stationarity.

A number of filters can also be applied to identify the cyclical component. Two most common time series filters are low-pass and high-pass filters. A low-pass filter allows low frequencies to pass the filter and subsequently removes all variance above the cutoff frequency. A high-pass filter analogously allows high frequencies to pass and removes all variance bellow the cutoff frequency. By combining the two a band-pass filter is created, which sets both an upper and lower cutoff frequency.

For identifying the cyclical component a majority of studies have identified the cyclical component as an ideal high-pass or an ideal band-pass filter. The supposed ideal high-pass filter was originaly proposed by Whittaker (1923) and popularized by Hodrick and Prescott (1980; 1997) which returns the sought for cycle component with the wanted duration by adjusting the smoothing parameter. The Hodrick-Prescott filter is a popular filter used in macroeconomic analysis (see, e.g., Danthine and Girardin (1989), Kydland and Prescott (1990)). Despite its wide usage economists are aware of its shortcomings, as has been identified by Harvey and Jaeger (1993), Cogley and Nason (1995) and Park (1996) for including spurious cycles in the filtered time series. Hamilton (2017) argues that the Hodrick-Prescott (HP) filter should not be used as it involves several levels of differencing. Therefore, the patterns observed in the data are likely to be the results of the filter itself and not the original data. Instead he proposes a regression filter that extracts the cyclical behavior with greater precision, but Schüler (2019) shows that it emphasizes frequencies longer than a typical business cycle and completely mutes the short frequencies, which is a major drawback when analyzing drought effects.

Baxter and King (1995) have established the approximate band-pass filter with changeable number of lags incorporated in the computing process. Similarly to the HP filter, it extracts cycles of length between 1.5 and eight years. Pedersen (2003) argues that this band-pass filter is generally superior to a band-pass filter based on the HP filter.

Bandi et al. (2019) use a simple low-pass filter in the time domain to calculate the frequency-specific components of a covariance-stationary bivariate process. It is characterized as being nonparametric in nature, because it hinges solely on suitably selected moving averages of the available data. It allows the user to modify the length of frequencies that are selected by modifying the range of values from which the moving average is constructed. The components of the bivariate process are the that are subject to the low-pass filter are returns of various portfolios. Therefore, the filter being applied to returns of commodities should not pose troublesome behavior in the form of extracting spurious cycles, leaving important information about the cyclical component present in the series.

2.2 Haar Wavelet Transform

The choice of filter is not accidental by Bandi et al. (2019), who observe that "any factor for which a classical Wold representation applies (Wold, 1938) always admits an equivalent orthogonal representation in which the factor can be modeled as an infinite sum of orthogonal frequency-specific components"(p.2). Their filter of choice is generally referred to as a Haar filter, because it uses the Haar scaling function to smooth the data. The Haar Wavelet Transform (HWT) decomposes the series in this manner and allow for exact reversibility of the orthogonal components without edge effects commonly visible with the other more complex wavelet algorithms, such as the Daubechies wavelet or Mexican hat wavelet. The Haar mother wavelet uses a rectangular window of length 2^j, where $j \in \mathbb{N}_{\not 0}$, and grows until it fits the entire series. The filter calculates moving averages by shifting the window over the input data and does so for each of the specified $j \in \mathbb{N}_{\not 0}$. In addition to this the Haar wavelet, due to being a simple concept, is very fast to compute and works well with financial time series. Kaplan (2001) argues that the higher resolution for smoothly changing the time series by using the more complex wavelets is not worth the cost for financial time series, because of jagged transitions within the series.

2.3 Spectral factor models

Traditional linear factor models such as the capital asset pricing model (CAPM) by Sharpe (1964) and Lintner (1965), the arbitrage pricing theory (see Ross (1976)), or the three factor model by Fama and French (1993) are commonly used to estimate returns, while calculating the risk associated with the given asset. The three factor Fama-French model captures the impact of the factors, the risk premium, size (i.e., market capitalization) and the book-to-market ratio of the portfolio on the returns of the examined portfolio. The coefficients (betas) presented with the individual variables in the three factor model illustrate the risk associated with the portfolio. According to Bandi et al. (2019) traditional factor models are viewed as restricted models to the spectral factor models in the way that they measure equal risk across all frequencies, meaning that the beta stays constant. This is a great limitation as it provides no information about which investment horizon best suits the portfolio. This constraint is eliminated with the spectral models, because they allow for a varying beta and therefore for frequency specific risk. Bandi et al. (2019) show that the traditional betas observable in the Fama-French model are weighted averages of the spectral betas, in which the weights reflect the information content (relative variance) of the individual frequencies.

The analysis approach of the time-frequency domain through spectral factor models is not the only option. A Fourier transform introduced by Fourier (1822) decomposes a signal into a linear combination of sine and cosine functions at different frequencies, which can then be summed together to obtain the original function. The drawback of the Fourier transform is the loss of all time domain information. Gabor (1946) created a short Fourier transform applied to a moving window throughout the series to retain some level of time information. It is nevertheless limited to a single pre-selected frequency. Therefore, to capture the informational content of both the time and frequency domain, wavelet analysis was developed and presents itself as a viable alternative to the spectral factor models. The ability to control the environment in such manner has lead to extensive research in all fields dealing with signal-like functions. It has also been utilized to provide a better insight into the behavior of commodity prices. Davidson et al. (1997) uses the wavelet analysis to extract information about commodity prices by constructing 'patios' to graph their behavior at multiple time scales across a range of frequencies. Other studies use it to examine co-movements of stock markets and commodities (see Vavřina (2012)), long memory in commodity futures volatility (see Elder and Jin (2007)), or the potential increase in predictive power as proposed by Davidson et al. (1997) and Ramsey (1999). Kriechbaumer et al. (2014) uses the wavelet analysis in the combination with ARIMA models and demonstrates an improved performance in forecasting metal prices.

However, Bandi et al. (2019) argue that the spectral factor models used in their work have both econometrical and economical advantages. Econometrically, they are explicit about the data generating process and the properties of the filter. With the betas being well defined and with the presence of a model, the spectral factor model outperforms the wavelet analysis. The spectral analysis is able to reproduce a similar outcome by the application of a filter to the data. The combination of filters at different time scales leads to a comparable result, but additional information, which leads to a better interpretation. Also, the spectral factor model with different spectral betas across frequencies implies delays in the adjustment of asset prices to changes in the factors, which is not possible using wavelet filters.

2.4 Drought indices

The thesis focuses on the impact of drought on commodity prices. The commodity prices do not leave much space for representation as they depict a recorded value that was present in the market at a given time. There is not, however, a general consensus about the approach of measuring drought. A wide range of indices have been developed over the years of researchers quantifying drought. An appropriate drought index must be chosen for the analysis to produce viable results. All of the presented indices provide a global overview of drought across time.

2.4.1 Palmer drought severity index

The PDSI was a milestone for drought indices as it enabled researchers to measure both the dryness and wetness on a single scale (see Palmer (1965)). It has since been refined by Wells et al. (2004), who introduced the self-calibrated PDSI (sc-PDSI), which improved the spacial comparability, but still maintained the shortcomings of a fixed temporal scale (between 9 12 months) and a lengthy autoregressive characteristic (see Guttman (1998)).

2.4.2 Standardized precipitation index

McKee et al. (1993) introduced the Standardised Precipitation Index (SPI). The common conception of drought being a multi-scalar phenomenon led to the creation of this index. This allows the inclusion of the time scale over which water deficits accumulate into the calculation, and thus enables to differentiate between types of drought (e.g., meteorological, agricultural, climatological, etc.).

2.4.3 Standardized precipitation evaporation index

Vicente-Serrano et al. (2010) argue that the inclusion of temperature into the calculation of drought (included in sc-PDSI but not in SPI) is vital when measuring drought, and show evidence of rising temperatures markedly affecting the severity of drought. Therefore, they construct the SPEI which maintains the feature of SPI of being multi-scalar adding potential evapotranspiration (PET) to the equation. The difference between the precipitation and evapotranspiration is taken and standardized to obtain the best possible value for the index (for specifics on the calculations see Vicente-Serrano et al. (2010)). SPEI uses two different methods to compute potential evapotranspiration (PET). One utilizes the Thorntwaite (1948) equation, which only uses data for temperature and latitude to compute PET. Varouchakis and Corzo (2019) use the Global Drought monitor and the Thorntwaite equation to analyze extreme hydrological events across the world. This method is used in several studies focusing on water deficits and potential evapotranspiration (see, e.g., Dragotă et al. (2011), Al-Sudani (2019) or Calvo (1986)). The second method uses the Penman-Montheith equation introduced by Allen et al. (1998), which is considered a superior approach as it utilizes a wider variety of input data. Zhou and Hong (2013) use this method for climate change projection and show the benefit of better quality forecasts.

2.4.4 Choosing time scale for SPEI

SPEI is calculated at different time scales, depending on how many lags are included to generate the index. The past data has the same weight as the data for the current month, when calculating the index for various time scales.

SPEI-01 only uses data for the given month and represents short-term soil moisture and crop stress, whereas SPEI-03 (3 month lag) reflects short- and medium-term soil moisture and a seasonal estimation of precipitation (for specifics on individual time scales see World Meteorological Organization (2012))

Agriculture drought is generally observed at a $3-6$ month time scale (see, e.g., Potopová et al. (2013)). Hudson (2019) found SPEI-06 (6 month lag) to have the greatest impact on crop yield in the Northeast U.S.. Bachmair et al. (2018) show that SPEI with short periods of water accumulations (around SPEI-03) are best linked to Vegetation stress indices, implying the index at this time scale has the greatest impact on crops in various places in Europe.

Zipper et al. (2016) identifies short term ($1 - 3$ month) drought time scales as having the largest impact on crop production of maize and soybeans in the U.S.. The inclusion of time lags in the index may cause unwanted distortions, such as autocorrelation of the error term. SPEI-03 may also be a problem, because it itself is already calculated over a certain horizon and the filter would then recalculate the series to fit the wanted horizon for the spectral factor model. This would cause complications with the filter setting, as it would try to change the already established timescale. Therefore, even though research suggests that agricultural drought appears at a $3 - 6$ month lag, SPEI-01 will be used in the regressions. As proved by Bandi et al. (2019) the wavelet filter applied to the data prior to regressing incorporates lags into the moving average. The agricultural drought time scales correspond to the high frequency scales as they represent similar lengths (for specifics see Section 4.3). However, the time lags are included in different stages of the calculation, so a difference between the values is to be expected.

2.5 Forecasting

Researchers use many different methods for time series forecasting. The most simple one is an autoregressive (AR) model, which generates an estimate based on previous values and a stochastic error term. It represents an AR(p) process, where p is the amount of lags incorporated in the model. A moving average (MA) model represents the estimate as a linear combination of past residuals, with q in MA(q) being the number of lags. The combination of the two, with the addition of the level of differencing needed for the data to become stationary, creates an autoregressive integrated moving average (ARIMA) model. However, the data entering the model will already be stationary, therefore the level of integration will automatically set to zero. Seasonality can also be added to the model, but due to the modification of the series discussed in Section 4.3, it is unclear how to account for seasons while generating the model. The ARIMA(p, d, q) uses only univariate data, which will be sufficient for the benchmark forecasting model. However, the goal is to improve this standard with the research on the impact of a drought index on the forecasting commodity prices preceding the forecasting. The inclusion of the exogenous variable into the model creates an ARIMAX. Comparing results obtained from the two forecasting models generates mixed conclusions, as the independent variables may not contribute to the forecasting power but rather add unwanted error to the predictions (see, e.g., Ďurka and Pastoreková (2012), Ababio (2012), Vagropoulos et al. (2016) or Kongcharoen and Kruangpradit (2013)).

There are several methods of evaluating the accuracy of the forecasting model. Some of the most commonly used measures are the root mean square error (RMSE) and the mean average error (MAE). RMSE gives bigger importance to the highest errors, and is therefore sensitive to outliers. MAE shows the absolute deviation from the true values. Both of these indicators are very often utilized for evaluate models forecasting financial markets (see, e.g., Mallikarjuna and Rao (2019) or Pradesh et al. (2018)). Another widely used measure of forecasting models is the mean average percentage error (MAPE). However, the calculation of MAPE requires a division by the original value, which could turn out to be problematic as the series consists of values fluctuating around zero. Also as shown by Hyndman and Koehler (2006), if any of the true values are very close to zero, which due to the distribution of the series is true for several observations, the corresponding absolute percentage errors will be extremely high, thus hindering the information provided by MAPE. As the predicted values are the percentage changes to commodity prices, an important indicator is the number of times the model predicts the correct sign of the forecasted values, also known as a success ratio (SR) (see, e.g., Panchal and Patel (2017)). This will demonstrate the ability of the model to correctly anticipate a rise or fall in the price and offer a different insight to the performance of the model.

Chapter 3

Data description

3.1 Commodity price data

The commodity price data is taken from the Pink Sheet obtained in the World Bank (2020) data catalog[1], which contains monthly data for all the essential commodities investigated in this thesis. The values are given in nominal U.S. dollars and the monthly series contains records of prices dating back to 1960. The commodity prices are aggregated as the average value of the commodity for the given time period (month). The commodities in question are Wheat HRW (US), Cocoa (ICCO), Soybeans (US), Maize (US), Rice, 5% broken (Thailand), Sorghum (US), Barley (US).

3.2 Drought data

The drought index will be taken from the SPEI Global Drought Monitor[2] (created by Vicente-Serrano et al. (2010)), which uses the Thornthwaite equation to compute PET. This particular index is chosen because it has many advantages to the other indices, such as PDSI or SPI. It is very easy to work with and it contains data to the present day (not the case with SPEI database, which uses the Penman-Moneith method of calculating PET). It is an index that estimate the levels of drought globally and is commonly used for spatiotemporal analysis regarding drought in various places across the world, such as in Czech Republic by Potopová et al. (2015) Australia by Deo and Şahin (2015) or in Jordan by Törnros and Menzel (2014). The time scale chosen is the SPEI-01, which does

[1] Avialable from https://www.worldbank.org/en/research/commodity-markets, accessed on 10.02.2020.

[2] Available from https://spei.csic.es, accessed on 10.02.2020.

not incorporate any lags in the calculation, in order to prevent a collision of two methods of horizon extraction. SPEI has values ranging from below (-2) (extreme drought) to values above 2 (extremely wet).

Regions that are heavily involved in agriculture, specifically the agriculture commodities this thesis is interested in, will be extracted and the average value of the drought index will be taken as a time series data set. The drought data are taken from specific locations to prevent data collected in regions that have no association with growing the individual crops to affect the prices. Most of the data used is from the United States as they are considered a giant when it comes to producing agriculture commodities. The locations, estimations of the magnitude of crop production in the given area, and specifications of how prices in the sheet are acquired are summarized in the Table 3.1:

Table 3.1: Description of the commodity data

Commodity	**Area of drought data**	**Price origin**
Barley ($/mt)	The U.S. states Idaho, Montana and North Dakota (over 73% of production)	Spot, 20 days To-Arrive
Cocoa ($/kg)	The region western Africa (Côte d'Ivoire, Nigeria, Ghana, etc.), accounts for approximately two thirds of world production (see ICCO (1999))	International Cocoa Organization daily price, average of the first three positions on the terminal markets of New York and London, nearest three future trading months
Maize ($/mt)	Midwest in the USA (over 60% of production).	f.o.b. US Gulf ports
Rice, Thailand 5% ($/mt)	Grown across the entirety of Thailand therefore drought data for the whole country has been taken	Indicative price based on weekly surveys of export transactions, government standard, f.o.b. Bangkok
Sorghum ($/mt)	US states Kansas, Oklahoma and Texas (over 80% of production)	Texas export bids for grain delivered to export elevators, rail-truck, f.o.b. Gulf ports
Soybeans ($/mt)	Midwest of the U.S. (approx. 70% of production)	Nearest forward
Wheat, US HRW ($/mt)	Spread out across the U.S. but majority concentrated in Great plains (over 50% of production)	Export price delivered at the US Gulf port for prompt or 30 days shipment

Chapter 4

Methodology

4.1 Introduction

The traditional factor models, such as the Fama and French (1993) three factor model, have company-specific features and analyze their riskiness to compute the estimated returns on the stocks of the company. Alternatively, a factor model for commodity futures prices can be implemented, which uses components, such as spot prices or convenience yield (see, e.g., Schwartz and Smith (2000)). However, in this thesis the aim is to find the impact of drought on commodity prices. Therefore, drought-specific features will be used to evaluate the risk associated with returns on commodities. The analysis will be focusing on the coefficients generated by the Ordinary Least Squares (OLS) regression, which can be interpreted as risk associated with the explanatory variable, in this case drought. Information about specific frequencies (cycles) is not available through analysis of a traditional factor model by itself, as it would generate results merely for one frequency and restrain the additional information offered by the data. Therefore, the spectral factor models proposed by Bandi et al. (2019) will be used to obtain the frequency-specific information about the risk of drought affecting prices. A filter will be applied to the data prior to utilizing the OLS regression to restrict the fluctuations to a known range of frequencies to capture the frequency-specific spectral betas.

The model is specified as follows,

$$R_{i,t}^{(j)} = \alpha + \beta_i^{(j)} \times SPEI_{i,t}^{(j)} + \varepsilon_{i,t},$$

where the dependent variable R is the return on the commodity over time t, the independent variable $SPEI$ is the drought index over time t, α refers to the intercept (constant), ε represents the random error in the data, i marks the specific commodity in question and j determines the scale (frequency). This model will be estimated for each commodity at all studied frequencies to determine the effects of drought on commodity prices over different horizons.

4.2 Stationarity

For the spectral models to work, covariance-stationary data is needed for both the dependent and independent variables. The wavelet filter itself does not require the input data to be stationary, however, the filtered series will be plugged into an OLS regression, which requires this condition in order to avoid generating spurious results. The method to determine whether the time series is stationary is an Augmented Dickey-Fuller test (ADF). This will test the presence of a unit root, which will be (if found) consequentially removed to obtain a covariance-stationary series and thus get the best possible outcome. Note that the ADF test will be applied only after the relative changes (i.e., returns) of the commodity prices are taken. Testing the commodity price series prior to taking this measure would most probably detect a unit root as financial data is usually considered non-stationary. As mentioned in Section 2.1, the test for stationarity did not detect a unit root for the relative changes of commodity prices. After removing a quadratic trend from the SPEI-01 series, the ADF test did reject the presence of a unit root and thus confirm stationarity of the data.

4.3 Beta decomposition

The filter applied to the data is a low-pass filter based on a moving average. Moving averages $\pi_t^{(j)}$ of length 2^j are constructed from the time series $\{x_t\}_{t\in\mathbb{Z}}$, i.e.,

$$\widehat{\pi}_t^{(j)} = \frac{1}{2^j}\sum_{p=0}^{2^j-1} x_{t-p}.$$

The difference between two subsequential moving averages of lengths 2^{j-1} and 2^j is denoted as $\widehat{x}_t^{(j)}$, i.e.,

$$\widehat{x}_t^{(j)} = \widehat{\pi}_t^{(j-1)} - \widehat{\pi}_t^{(j)}.$$

The difference $\widehat{x}_t^{(j)}$ of the moving average for periods of lengths 2^{j-1} and 2^j

captures fluctuations that survive the averaging over 2^{j-1} but disappear for the average of 2^j values. This decomposes the time series into a set of bands. For j = 1 the highest frequencies of fluctuations are captured and with higher values of j the band reflects lower and lower frequencies, i.e., higher scale allows to extract longer cycles form the data. These frequency-specific components are referred to as discrete Haar transforms (DHT) of the original process. However, an infinite number of scales is unwanted and not even possible due to the length of the data and the window of the moving average. Therefore, the scales will be restricted to $j = 1, ..., 6$, and all lower frequencies will be grouped under $j > 6$. The frequencies are expressed in calendar time (months). The exact resolutions of the filtered data for different values of j are described in Table 4.1.

Table 4.1: Interpretation of the scales

Scale	**Frequency resolution**
$j = 1$	$1 - 2$ months
$j = 2$	$2 - 4$ months
$j = 3$	$4 -$ 8 months
$j = 4$	$8 - 16$ months
$j = 5$	$16 - 32$ months
$j = 6$	$32 - 64$ months
$j > 6$	> 64 months

After filtering the data, an initial observation is included multiple times in the new series. If the new series is not modified in any way, large serial correlation is observed for lower frequencies, which might hinder the interpretation of the regression model. Therefore every 2^{j-1}-th component will be selected from the data and the regression will be concluded on the reduced sample.

Because of the orthogonality of the decomposed components, Bandi et al. (2019) argue that the traditional beta is a weighted average of the frequency-specific spectral betas,

$$\widehat{\beta} = \sum_{j=1}^{J} \widehat{v}^{(j)} \widehat{\beta}^{(j)},$$

where J is the total number frequency specific components and $\widehat{v}^{(j)}$ is the weight (relative variance), i.e.,

$$\widehat{v}^{(j)} = \frac{\widehat{Var}(R_i^{(j)})}{\widehat{Var}(R_i)}.$$

This allows not only for an analysis of the magnitude of the spectral betas obtained from the regressions, but also for an evaluation of the weights of the coefficients at various frequencies. Meaningful information about how the individual spectral betas impact the traditional beta is stored in the relative variance, which will be calculated for each frequency.

4.4 OLS assumptions

As the commodity price series will be regressed on the SPEI drought index using OLS, the Gauss-Markov assumptions will have to be tested to obtain the best possible estimates of the frequency-specific spectral betas. Some level of bias should be expected, because there is only one explanatory variable (SPEI-01) used in the regression and many more would possibly have an impact on the commodity price returns (e.g., economic growth/recession, increased consumption, policies, etc.). However, none of these potentially impactful factors are added as control variables, in order to maintain simplicity of the spectral factor models. The evaluation of the individual assumptions is shown in Chapter 5. A problem commonly encountered in financial time series analysis is a non-normally distributed error term. Non-normality allows the model to generate an efficient estimator $\widehat{\beta}$, but it proves to be a major limitation because it does not allow for hypothesis testing, and subsequently reveals at which time scales the coefficients are significant.

Regression literature offers several solutions to achieve robust inference despite this difficulty. The easiest solution is using the Central Limit Theorem (CLT) which assumes the errors are normally distributed if the sample size reaches a high number of observations (approximately $n > 30$). Both of the series inspected in this thesis contain a high number of observations, but it appears that regressions at a number frequencies produce heavy tails to their distributions. An alternative to the parametric tests as t-tests or F-tests are rank based non-parametric methods, which do not make approximations of the distribution of the outcome. These tests generate p-values of the estimates by ranking the data, they are however not as powerful as the parametric tests, when the assumptions are met. A different approach to avoid the problem of non-normality is bootstrapping, which resamples the available data randomly with replacement. It makes the results of a regression robust by allowing to assign measures of accuracy (e.g., confidence intervals) to the estimates. Kang et al. (2012) base their bootstrap procedure on the stationary bootstrap of

Politis and Romano (1994) to obtain p-values for the timescale betas. Bandi et al. (2019) use a robust inference method proposed by Ledoit and Wolf (2008; 2011), which is also based on the stationary bootstrap.

4.5 Bootstrap procedure

The bootstrap used to resample the data, leading to robust inference of the estimates, is based on the stationary bootstrap of Politis and Romano (1994). This procedure is a block bootstrap. The individual block lengths are distributed as a geometric random variable and the optimal length is chosen by an algorithm used by Politis and White (2004) and later corrected by Patton et al. (2009). The bootstrap procedure will generate 10000 bootstrapped regressions, from which the desired indicators will be derived. The p-values are obtained by using the method of Ledoit and Wolf (2008; 2011). The adjusted R^2 will be taken as the estimated value from the bootstrapped distribution and confidence intervals will be created. According to DiCiccio and Efron (1996), bootsrapping generates asymptotically better confidence intervals than the ones obtained by using sample variance with the assumption of normality. Therefore, confidence intervals for R^2 will be based on the percentiles of the distributions, i.e., plotting the 2.5th and 97.5th percentiles to obtain a 95% confidence interval (see, e.g., Kang et al. (2012)), rather than using the normality assumption and the R^2's significantly different from zero will be highlighted.

4.6 ARIMAX

Similarly to the OLS the input data will be reduced to $\frac{1}{2^{j-1}}$ of the sample. The frequencies at which the beta coefficients are significant will be used in the forecasting model as they can reveal useful information for the prediction. The ARIMA model will be considered superior to the ARIMAX model at frequencies in which the drought index does not prove to have a significant impact on the commodity prices. The inclusion of the exogenous variable might improve the forecasting abilities but it would likely be a statistical fluke as no causal relationship between the two variables was found.

Seasonality will not be included in the forecasting models as the applied filter and reduction of the dataset blend seasons together, resulting in vague characterization of a season.

The series is divided into two subsets, a training set and a test set. An algorithm evaluates the best possible parameters for the ARIMA model by evaluating the Akaike Information Criterion (AIC) for the training set. Note that the model, which achieves the lowest AIC, is found for each frequency separately. After the best possible model is chosen, an in-sample forecast of length of the test set will be made, which will be then be used to evaluate performance. As the aim of the thesis is to compare, which of the two models produce better forecasts, an out-of-sample forecast is not desirable as there is no data to compare it to. The length of the forecast will differ for various frequencies (see Table 4.2). Otherwise a short forecast for the low frequencies would require a relatively much longer forecast for the high frequency data. Due to this, the lengths of the training and testing sets will differ for individual frequencies.

Table 4.2: Interpretation of the scales for forecasting periods

Scale	**Forecast period**
$j = 1$	1 year
$j = 2$	2 years
$j = 3$	4 years
$j = 4$	8 years
$j = 5$	12 years
$j = 6$	18 years
$j > 6$	20 years

Chapter 5

Results and discussion

This sections summarizes the results of the empirical analysis and draws possible conclusions. Firstly, in order to generate an efficient estimator $\widehat{\beta}$, the OLS classical assumptions need to be verified.

The model is correctly specified, it has an error term which contains the variation of the dependent variable unexplained by the model and is therefore **linear in parameters**.

The mean of the error term ε is zero, because of the inclusion of the constant α in the model, i.e., any non-zero mean of the error term is absorbed by the constant (both of the series have a mean zero, so the intercept is usually a value very close to zero or statistically insignificant).

The assumption of **all explanatory variables being uncorrelated with the error term** is tested using the Pearson correlation coefficient test statistic. The results indicate that the there is no correlation between the drought index and the error term, but a rather high correlation with the dependent variable, due to only one independent variable included in the model.

The assumption requiring **no perfect multicollinearity** between explanatory variables is automatically satisfied as there is only one explanatory variable in the model.

To test **heteroskedasticity** of errors (i.e., constant variance of errors over time), the Breush-Pagan test will be utilized. The results are mixed for different frequencies. For some the null hypothesis of homoskedasticity is rejected but for other frequencies there is a lack of evidence for rejecting it. The heteroskedasticity consistent HC3 error are used for all regressions as Long and Ervin (2000) found HC3 to have superior performance over the other robust errors in small and large samples. They also showed that it is not inferior to the OLS standard error even when no heteroskedasticity was exhibited by the model.

The error term must not be **autocorrelated** in order for the model to generate unbiased estimators and to be able to utilize the hypothesis tests to find out at which frequencies are the spectral betas statistically significant. To reveal autocorrelation in the error term the Durbin-Watson test is used. The results differ across frequencies. Autocorrelation is rejected for all commodities when no filter is applied (regressing the original series) and at most frequencies. For frequencies where serial correlation is present, the Newey-West standard errors will be used as they correct for both autocorrelation and heteroskedasticity.

The **normality of the error term** is not satisfied at most frequencies if all data points are used in the regression. However, after a reduction of the sample to $\frac{1}{2^{j-1}}$ of the size by taking only every 2^{j-1}-th component, the occurrence of normality of residuals has considerably increased, but is still non-normal for some frequencies. This could be given by the fact that the model consists of only one explanatory variable, and other unidentified factors affecting the commodity price returns are not included. Therefore, to achieve robust inference of the coefficients, the bootstrapping procedure discussed in the Section 4.5 is used.

5.1 Spectral factor models

The regression results are visible in the Table 5.1. Values for the beta estimates, p-values, adjusted R^2 and the relative variance (weight) of each frequency-specific component are recorded for every commodity. Significance levels for the beta estimates and adjusted R^2 are marked with an asterisks as mentioned in the table description. The values of the adjusted R^2 in the tables represent the estimated value from the bootstrap procedure. In Figure 5.1, beta estimates for all commodities are shown for particular scales, with the significant and insignificant coefficients differentiated by color.

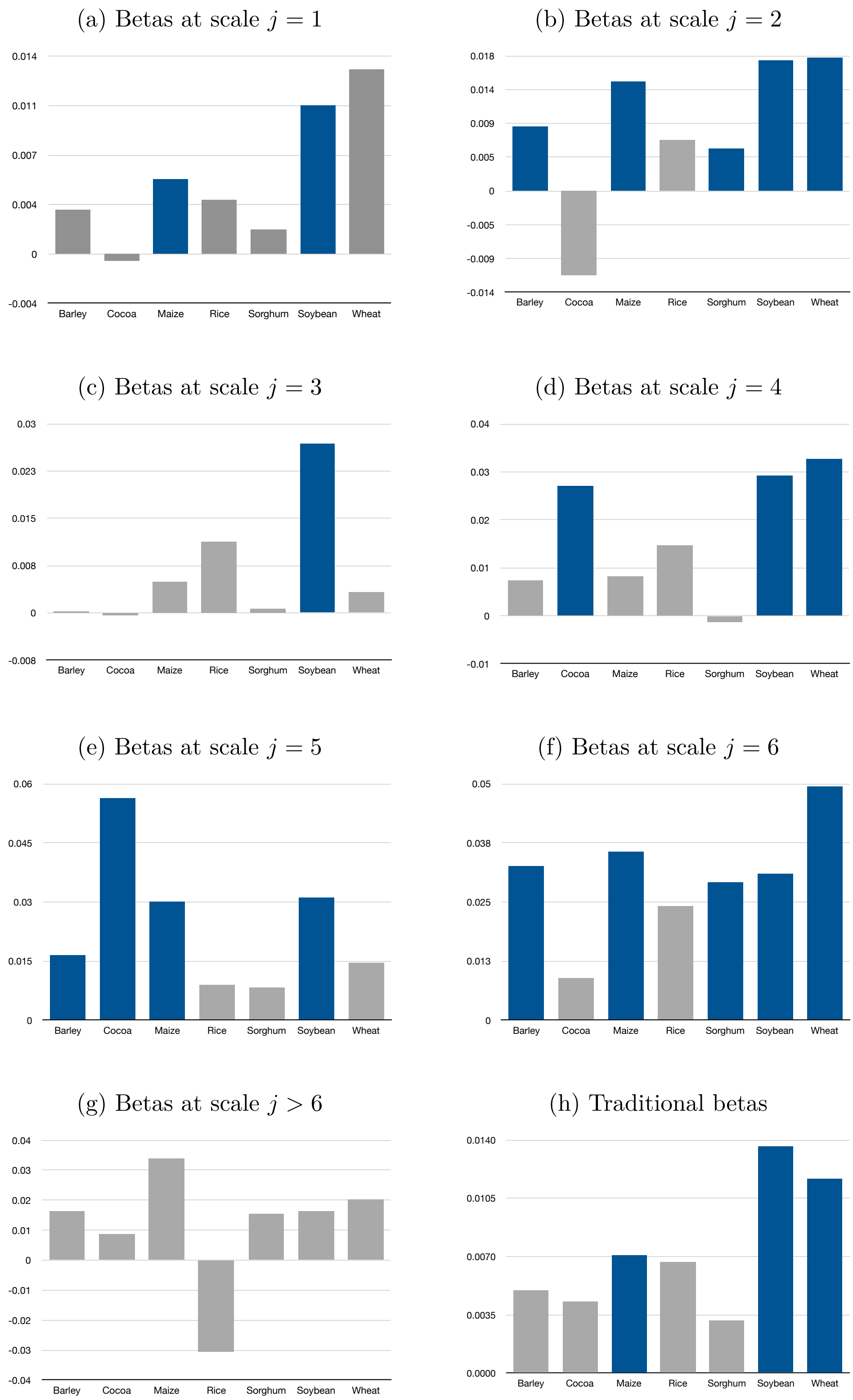
Note that the variance of SPEI-01 is approximately 100 times higher than the one for the relative changes in commodity prices. Therefore, the values of the beta coefficients appear to be very low. Due to SPEI representing the occurrence of drought with negative values, the index was multiplied by (-1), to achieve positive coefficients for a positive relation.

Table 5.1: Beta decomposition

Commodity	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$	$j > 6$	$\sum_{j=1}^{7} v^{(j)} \widehat{\beta}_i^{(j)}$
Barley								
$\widehat{\beta}$	0.003	0.009*	0.000	0.008	0.017*	0.033**	0.016	0.005
$(p - value)$	(0.498)	(0.062)	(0.956)	(0.416)	(0.064)	(0.048)	(0.491)	(0.237)
Adj. R^2	0.002	0.006	−0.003	0.011	0.028	0.104	0.078	0.003
weight	0.448	0.232	0.135	0.091	0.058	0.027	0.010	
Cocoa								
$\widehat{\beta}$	0.000	−0.011	−0.001	0.027***	0.056**	0.009	0.009	0.004
$(p - value)$	(0.924)	(0.149)	(0.926)	(0.010)	(0.016)	(0.749)	(0.810)	(0.496)
Adj. R^2	0.000	0.006	−0.003	0.037	0.122*	0.014	−0.054	0.001
weight	0.353	0.225	0.198	0.101	0.051	0.021	0.026	0.026
Maize								
$\widehat{\beta}$	0.005**	0.015***	0.005	0.008	0.030*	0.036**	0.034	0.007*
$(p - value)$	(0.031)	(0.001)	(0.610)	(0.273)	(0.096)	(0.035)	(0.432)	(0.068)
Adj. R^2	0.004	0.020***	0.007	0.007	0.104	0.178	0.245	0.006
weight	0.455	0.225	0.126	0.088	0.060	0.033	0.012	
Rice								
$\widehat{\beta}$	0.004	0.007	0.011	0.015	0.009	0.024	−0.031	0.007
$(p - value)$	(0.361)	(0.339)	(0.109)	(0.168)	(0.129)	(0.144)	(0.817)	(0.130)
Adj. R^2	0.004	0.007	0.016	0.038	−0.009	0.034	−0.019	0.008
weight	0.386	0.223	0.175	0.110	0.078	0.022	0.006	
Sorghum								
$\widehat{\beta}$	0.002	0.006*	0.001	−0.001	0.008	0.029***	0.016	0.003
$(p - value)$	(0.652)	(0.084)	(0.875)	(0.807)	(0.346)	(0.001)	(0.552)	(0.278)
Adj. R^2	0.000	0.002	0.002	−0.003	0.002	0.123	0.037	0.001
weight	0.369	0.225	0.165	0.128	0.069	0.033	0.011	
Soybeans								
$\widehat{\beta}$	0.011***	0.017***	0.027***	0.029***	0.031***	0.031***	0.016	0.014***
$(p - value)$	(0.004)	(0.001)	(0.003)	(0.003)	(0.004)	(0.009)	(0.559)	(0.002)
Adj. R^2	0.012**	0.021**	0.039***	0.043**	0.065	0.074	0.182	0.018***
weight	0.451	0.228	0.125	0.085	0.061	0.033	0.013	
Wheat								
$\widehat{\beta}$	0.013	0.018***	0.003	0.033*	0.015	0.049***	0.020	0.012**
$(p - value)$	(0.128)	(0.005)	(0.742)	(0.091)	(0.234)	(0.005)	(0.482)	(0.014)
Adj. R^2	0.019	0.018**	0.001	0.073	0.006	0.234*	0.074	0.012*
weight	0.447	0.218	0.130	0.091	0.063	0.039	0.012	

* p<0.10, ** p<0.05, *** p<0.01

Figure 5.1: **Frequency-specific betas.** Blue columns indicate significant betas at a 90% significance level, gray columns indicate insignificant betas.



5.1.1 Trends in $\widehat{\beta}$, R^2

A clear trend can be observed, when examining the magnitude of the frequency specific coefficients. The values are higher when the frequencies are lower (i.e., the scale j is higher). This is not necessarily the case for all betas but is apparent for the significant ones. The same behavior is exhibited by the adjusted R^2, where at lower frequencies SPEI-01 describes the price changes better than at the high frequencies. This phenomenon is likely to be a consequence of the filter applied.

Firstly, the weight of the spectral betas exhibits a downward sloping trend unlike the betas or the adjusted R^2's. The relative variance becomes lower every time the scale j grows, leaving only about 1% of the total variance for the lowest frequency component, similar to the findings of Bandi et al. (2019). At high frequencies the data resembles noise, but by applying the filter, the series is smoothed out, starts to display longer fluctuations (cycles), and would eventually converge to a constant, if the scale were high enough. In this extreme case the model would be explained perfectly, so the model is able to analyze the long cycles much easier, unlike if the variables contain a high amount of abrupt changes.

Secondly, the length of the series is reduced to $\frac{1}{2^{j-1}}$-th of the original sample (see Section 4.3). The model therefore does not have as many data points available and assigns a higher weight to each individual point, thus increasing their impact.

It is also noticeable, that a relatively low number of adjusted R^2's are deemed statistically significant after creating confidence intervals from the bootstrap distribution. This is probably caused by the values being relatively low to begin with due to the inclusion of only one explanatory variable.

5.1.2 Interpretation of the $\widehat{\beta}$

A statistically significant impact of SPEI-01 on commodity prices, when considering the traditional factor model, i.e., the unfiltered series, has been found for three commodities (maize, soybeans and wheat, US HRW). This means that the prices of the commodities are subject to risk associated with drought regardless of the frequency. Generally, if the regression revealed statistically significant relationships at a high number of frequencies, the chance of the overall relationship being significant is higher. All of the three aforementioned commodities also have a statistically significant beta at a 99% level at a high frequency. Due to the high weight of the components at high frequencies, it greatly increases the chance of the traditional beta being significant.

The explanatory power of the model concerning maize is rather questionable as the estimated R^2 from the bootstrap is just 0.006. The regressions regarding soybeans and wheat showed greater impact with significant values of both the coefficient and the R^2.

For most of the commodities, a significant element of risk of drought affecting the commodity prices has been found for at least some frequencies.

Rice, Thai 5% is the only commodity from the list, for which the relationship of SPEI-01 and commodity price has not been found at a sufficient confidence level for any frequency, with the lowest p-values being just above 0.1. This may be caused by the generalization of a large area, causing the effects of SPEI-01 to negate themselves, which is mentioned by Prabnakorn et al. (2018), who studied the correlation of SPEI-01 with rice yield in Thailand on a much smaller spacial scale.

Regressions for soybeans have revealed a significant relationship between the dependent and independent variables at all frequencies apart from the lowest frequency resolution ($j > 6$). This shows that the soybean prices are at risk from being affected by drought, when looking at short and long cycles up to 64 months. Again the magnitude of the impact of drought is higher when considering long cycles in comparison with the short-term fluctuations.

The individual frequencies at which most of the commodity prices experience a significant impact of SPEI-01 are the second and the sixth scale ($j = 2, 6$). Five out of the seven commodities in question exhibit this behavior. No significant impact of drought was found at the lowest frequency scale ($j > 6$), where the cycles exceed 64 months in length. The frequency with only one commodity (soybeans) significantly impacted by SPEI-01 is represented by the third scale ($j = 3$).

The length of fluctuations measured at the scale $j = 2$ is equivalent to $2-4$ months. The high number of significant coefficients is a reasonable result, when considering it is usually short term droughts which affect the crop yields as shown for maize and soybeans by Zipper et al. (2016). Also it is sensible to mention, that the planting dates for most of the commodities significant at this scale often span across approximately two months (see United States Department of Agriculture (1997). The growing season is usually around $4-5$ months (wheat takes longer as it is grown over the winter), which similarly to the planting dates falls into the range of cycles captured by the second frequency scale. Therefore, it is possible that the positive relationship between drought and commodity prices captured at high frequencies is a representation of crops being affected by drought during the growing seasons and the market prices adjusting accordingly.

The high number of commodities that revealed themselves to be at risk of SPEI-01 affecting prices for long cycles (scale $j = 6$) might not be a coincidence. Drought may not be the main catalyst of the rising commodity prices, but it often appears to contribute in times when commodity prices experience major increases in value. For example, Trostle (2011) shows the impact drought had on the major increase of commodity prices after the recession of 2008, Light and Shevlin (1998) attribute the price shock in U.S. grain prices partly to the drought in the Midwest. Major droughts were also recorded in the late 1980s in North America and caused a rise in prices of grains and other agriculture commodities (see, e.g., Riebsame et al. (1991)). With other synchronized co-movements of the two series, which were not mentioned, in addition to the cycles exposed by the application of the wavelet filter, the impact of drought on commodity prices in long cycles appears to be much more understandable.

Another possible explanation for the phenomenon of the commodity prices being affected in long cycles, is the fact that nominal prices were taken instead of real prices. Whereas real commodity prices have been experiencing a constant or even a declining trend, the nominal prices have been experiencing an upward trend. The difference between the two is expected due to inflation, but the individual changes may be more pronounced due to a higher inflation of the U.S. dollar (i.e., the currency in which the commodities are traded). Also because inflation is not a macroeconomic measure which is known to exhibit short-term fluctuations, the effects are instead more prevalent in the medium- to long-run.

To determine which frequency best represents the overall impact of SPEI-01 on the commodity price changes, the magnitude of the estimated coefficients must be examined. The size of the traditional beta, with values around 0.1, falls between the sizes of the two highest frequencies, the scales $j = 1$ and $j = 2$. However, the majority does not exhibit a significant relationship at the scale $j = 1$, therefore it can be concluded that the frequency scale $j = 2$ is best representative of the overall impact of SPEI-01 on commodity prices.

5.1.3 Agricultural drought

Agricultural drought is usually observed when $3 - 6$ months worth of lags are included in the calculation of SPEI. This corresponds to the second and third scale ($j = 2, 3$) utilized in the filter. From the Table 5.1 it is visible that the second scale ($j = 2$), along with the sixth scale ($j = 6$), have recorded the most statistically significant coefficients, which indicates an apparent relationship in these short frequencies. The third scale ($j = 3$) recorded only one significant beta and is thus unlikely to signify agriculture drought. To see, whether the second scale can be taken as a representation of the agricultural drought, the unfiltered commodity price changes are regressed on unfiltered SPEI-04 (in the range of agriculture drought and $j = 2$).

Table 5.2: Agriculture drought

Commodity	$\widehat{\beta}$	$(p - value)$	Adj. R^2	90% confidence interval of R^2
Barley	0.002	(0.510)	0.001	[-0.001, 0.007]
Cocoa	0.012**	(0.036)	0.007	[-0.001, 0.020]
Maize	0.002	(0.678)	0.001	[-0.001, 0.008]
Rice	0.008**	(0.044)	0.008	[-0.001, 0.024]
Sorghum	0.009***	(0.005)	0.011*	[0.001, 0.026]
Soybeans	0.007	(0.114)	0.006	[-0.001, 0.021]
Wheat	0.009**	(0.041)	0.007	[-0.001, 0.020]

* p<0.10, ** p<0.05, *** p<0.01

From the bootstrapped regression results summarized in Table 5.2 it is shown, that four out of the seven commodities are significantly impacted by SPEI-04 with a 95% confidence level. But only two (sorghum and wheat) of the four were also significant at the second frequency scale in the spectral factor model. Therefore it cannot be said with certainty that the scale $j = 2$ is representative of the agriculture drought having an impact on the commodity prices. It may rather be a representation of short term droughts appearing with a relatively high frequency. The occurrence of flash droughts (perhaps not always as extreme) could play a role as they are short in duration, develop rapidly, and often appear in growing season (see, e.g., Otkin et al. (2017)). Kuwayama (2019) shows that even a drought of only one week can be potentially very damaging to crop yields, which can lead to a change in price of the commodity. A possible influence to these short term cycles could also be a level of speculation as commodities are traded in a financial market. As Brooks et al. (2014)

shows, commodity prices respond to new information about drought and prices rise by a greater amount if larger areas are under extreme drought conditions.

A possible explanation for why cocoa and rice are found to be significantly impacted by agriculture drought may be the fact that these are the only two commodities included that are not grown in the U.S. (for specifics see Table 3.1). The two regions, countries of West Africa and Thailand, are classified as developing countries and their farmers do not receive the financial support from the government in such magnitude as in the U.S.. For the cocoa producers the support in the form of subsidies differs from country to country (see, e.g., Wessel and Quist-Wessel (2015)). The U.S. government heavily subsidizes the agriculture sector. The largest subsidy program is crop insurance provided by the U.S. Department of Agriculture with disbursements averaging around $9 billion annually in recent years according to the Congressional Budget Office (2017). It must be noted that the big agriculture businesses have major lobbying power over the U.S. government, which puts pressure on legislators to direct the expenditures towards agriculture (see, e.g., Bellemare and Carnes (2015) or Riedl (2002)). Irrigated agriculture accounts for the usage of approximately 80% of the consumptive water in the U.S. and contributes to half of the overall value of crop sales on just above a quarter of the harvested cropland (see United States Department of Agriculture (2019)). This mitigates the impact of weather conditions due to the effective irrigation system in place. Carr and Lockwood (2011) show that irrigation is rarely used in the African cocoa production as it remains too expensive for most farmers. In Thailand less than 20% of cropland is under irrigated conditions according to Kupkanchanakul (2000). Therefore, big farmers growing crops in the U.S. have a clear advantage over those in Africa or Thailand and might be less susceptible to agricultural drought.

5.2 Forecasting

Now that individual frequencies where SPEI-01 significantly impacts commodity prices are found, a comparison of forecasting accuracy of two different models will be made, with one focusing solely on the available commodity price data and the other containing the drought index as an exogenous variable in addition to the already mentioned series.

For each model the order of the ARIMA(p, d, q) is reported, where d is equal to zero for all models as the input data is stationary. Three indicators determining the functionality of the model are considered, the root mean square error (RMSE), mean absolute error (MAE), and a percentage of how many times the model predicted the correct sign of the forecasted values (the success ratio depicted by SR). A total of 23 forecasts have been made for each of the two models and the results are presented in the Table 5.3.

The order of the ARIMA(p, d, q) may differ for the two corresponding models, as the algorithm determining the lowest Akaike Information Criterion (AIC) was used for both models to reveal the best possible setting. One exception has been made and that is, if the AIC returned the order (0,1) for ARIMA an alternative model has been used, because the generated forecast was a constant value of zero, i.e., mean, which minimized SR. The model with a value of AIC closest to the initial one was chosen in these cases.

Table 5.4 presents a summary of the performance measures listed in Table 5.3. Smaller values of the errors and a higher values of the success ratio (SR) indicate better performance.

37 instances out of a total of 69, where ARIMAX outperformed ARIMA, were recorded. ARIMAX model had lower values for RMSE and MAE in the majority of times, but the two models outperformed each other the same amount of times, when it comes to the success ratio. If the number of times, where SR $> 50\%$, is taken for each model, ARIMAX has achieved this 16 times and ARIMA 15 times.

Also it must be noted that for some models the indicators had the exact same values, which works in favour of the ARIMAX model, as an addition of the exogenous variable does not lead to a drop in performance. If summed up, ARIMAX exhibited improved or equal performance on 46 occasions, compared to the 26, where ARIMA performed better.

For the scale $j = 2$ ARIMAX performed better only twice out of five times. This leads to a very unclear verdict, whether the inclusion of drought on the forecasting model helps with the predictions in this situation, where short cycles $(2 - 4$ months) are considered. This could be due to the fact that the series is very unstable (i.e., noisy) at this high frequency, and is therefore more difficult to forecast. However, at the scale $j = 6$, which recorded the same amount of significant coefficients across the spectral factor models, the ARIMAX showed smaller values of both errors for every commodity, with only one instance (soybeans) of $\mathrm{SR}_{ARIMA} > \mathrm{SR}_{ARIMAX}$. But even in this case the $\mathrm{SR}_{ARIMAX} = 75\%$, which is still a better result than a majority of other forecasts. Therefore, it can be said that, when forecasting long-term (18 years) cycles of commodity

Table 5.3: Forecasting results

Commodity	**Frequency**	**Model**	**Order**	**RMSE**	**MAE**	**SR**
Barley	$j = 2$	ARIMA	(2,0)	0.0285	0.0225	46.15%
		ARIMAX	(2,1)	0.0279	0.0220	69.23%
	$j = 5$	ARIMA	(2,1)	0.0143	0.0121	60.00%
		ARIMAX	(1,1)	0.0144	0.0122	90.00%
	$j = 6$	ARIMA	(0,2)	0.0120	0.0113	25.00%
		ARIMAX	(0,1)	0.0117	0.0111	87.50%
Cocoa	$j = 4$	ARIMA	(1,1)	0.0129	0.0104	53.85%
		ARIMAX	(0,1)	0.0137	0.0110	53.85%
	$j = 5$	ARIMA	(1,0)	0.0100	0.0091	70.00%
		ARIMAX	(1,0)	0.0124	0.0112	10.00%
Maize	$j = 1$	ARIMA	(1,1)	0.0397	0.0321	46.15%
		ARIMAX	(1,1)	0.0395	0.0321	53.85%
	$j = 2$	ARIMA	(2,1)	0.0359	0.0246	69.23%
		ARIMAX	(2,2)	0.0364	0.0255	46.15%
	$j = 5$	ARIMA	(2,2)	0.0110	0.0082	70.00%
		ARIMAX	(2,1)	0.0108	0.0086	60.00%
	$j = 6$	ARIMA	(2,0)	0.0098	0.0085	50.00%
		ARIMAX	(2,0)	0.0095	0.0083	50.00%
	$\sum_{j=1}^{7} v^{(j)} \widehat{\beta}_i^{(j)}$	ARIMA	(2,0)	0.0456	0.0348	50.00%
		ARIMAX	(1,0)	0.0461	0.0362	37.50%
Sorghum	$j = 2$	ARIMA	(2,3)	0.0204	0.0167	61.54%
		ARIMAX	(0,2)	0.0206	0.0168	46.15%
	$j = 6$	ARIMA	(1,0)	0.0083	0.0070	100.00%
		ARIMAX	(1,0)	0.0071	0.0060	100.00%
Soybeans	$j = 1$	ARIMA	(2,1)	0.0216	0.0164	61.54%
		ARIMAX	(2,1)	0.0209	0.0165	53.85%
	$j = 2$	ARIMA	(2,1)	0.0166	0.0137	76.92%
		ARIMAX	(2,1)	0.0188	0.0152	38.46%
	$j = 3$	ARIMA	(2,1)	0.0167	0.0125	53.85%
		ARIMAX	(0,1)	0.0148	0.0115	76.92%
	$j = 4$	ARIMA	(2,0)	0.0106	0.0091	69.23%
		ARIMAX	(2,0)	0.0092	0.0081	76.92%
	$j = 5$	ARIMA	(2,1)	0.0121	0.0082	80.00%
		ARIMAX	(2,2)	0.0109	0.0075	70.00%
	$j = 6$	ARIMA	(1,1)	0.0058	0.0053	87.50%
		ARIMAX	(0,1)	0.0054	0.0049	75.00%
	$\sum_{j=1}^{7} v^{(j)} \widehat{\beta}_i^{(j)}$	ARIMA	(2,1)	0.0300	0.0240	53.85%
		ARIMAX	(2,1)	0.0288	0.0221	61.54%
Wheat	$j = 2$	ARIMA	(1,1)	0.0283	0.0238	46.15%
		ARIMAX	(2,2)	0.0273	0.0233	69.23%
	$j = 4$	ARIMA	(1,1)	0.0164	0.0130	53.85%
		ARIMAX	(1,1)	0.0154	0.0124	38.46%
	$j = 6$	ARIMA	(1,1)	0.0147	0.0134	50.00%
		ARIMAX	(0,2)	0.0135	0.0120	87.50%
	$\sum_{j=1}^{7} v^{(j)} \widehat{\beta}_i^{(j)}$	ARIMA	(1,2)	0.0230	0.0179	46.15%
		ARIMAX	(0,1)	0.0230	0.0179	69.23%

Table 5.4: Summary of forecasting results

Performance	**RMSE**	**MAE**	**SR**	Sum
ARIMA > ARIMAX	7	9	10	26
ARIMA < ARIMAX	15	12	10	37
ARIMA = ARIMAX	1	2	3	6
Sum	23	23	23	

prices, it is beneficial to include a variable representing drought (SPEI-01), as the model displayed better results in nearly all the cases. The models at other scales experienced an improvement in performance, when drought was introduced, only occasionally, thus making the decision of which of the models is better to use very difficult.

Among the individual commodities barley, soybeans and wheat reported lower values for the error measures with SR > 50% for a majority of the forecasts, when the ARIMAX model was used. On the other hand, cocoa showed ARIMA performing better in both cases. All remaining commodities had mixed results of the indicators, leaving it difficult to determine which model performed better.

Chapter 6

Conclusion

This thesis analyzes the relationship of drought and commodity prices since 1960 and examines the cyclical effects of both elements. The first part of the analysis uses the spectral factor model framework proposed by Bandi et al. (2019), to find evidence of drought impacting commodity prices in a variety of frequencies. The second part focuses on the forcastability of the commodity prices and determines, whether the inclusion of drought to the model is relevant for the research. The analysis includes seven commodities, Wheat HRW (US), Cocoa (ICCO), Soybeans (US), Maize (US), Rice, 5% broken (Thailand), Sorghum (US), Barley (US), and a drought index SPEI-01.

The first goal of this thesis is to examine the frequency specific risk of drought affecting commodity prices. To find a relationship between the two variables, the series were examined at different frequencies, ranging from short ($1-2$ months) to long (more than 64 months) cycles. The Haar wavelet filter was utilized to decompose the traditional factors into frequency-specific components. The Haar wavelet filter allows for an exact reversibility of the of the orthogonal components obtained though the decomposition, which enabled an evaluation of the weights of individual frequency-specific betas. The weights revealed a decreasing trend, with the lowest frequency betas having only a minor impact on the traditional betas. However, with lower frequencies the spectral betas and the corresponding adjusted R^2's were greater in magnitude compared to the ones exhibited at higher frequencies. All commodities apart from rice, have shown a statistically significant risk of being affected by drought at some of the examined horizons. Five out of seven commodities have shown a statistically significant impact of SPEI-01 at frequency scales $j = 2, 6$, which translate to fluctuations of lengths $2-4$ months and $32-64$ months respectively.

The two commodities not impacted by SPEI-01 at the scales $j = 2, 6$ (rice and cocoa) were however found to be significantly impacted by SPEI-04, which was chosen as a representative time scale of agriculture drought. Coincidentally they are the only two commodities not grown in the U.S., which might be a portrayal of farmers in Thailand and countries of West Africa having a lower accessibility to irrigation or government support, compared to the agricultural sector in the U.S.. It remains unclear, why the U.S. grown commodities are significantly impacted at $j = 2$, however it may be the case of the crops in the U.S. being more susceptible to short-term droughts, resulting in crop stress, occurring in short cycles.

At frequencies, which revealed a statistically significant impact of SPEI-01 on the commodity prices, forecasting models were created. For each of the mentioned frequencies a best possible model with and without the inclusion of the drought variable was found using the Akaike Information Criterion (AIC) and their accuracy was compared using three indicators. These measures are the root mean square error (RMSE), mean absolute error (MAE), and a percentage of how many times the model predicted the correct sign of the forecasted values. The ARIMAX model outperformed the ARIMA model on more occasions, when evaluating the errors, and outperformed each other the same amount of times, when predicting the correct sign. ARIMAX exhibited superior forecasting accuracy for most commodities at frequency scale $j = 6$, for which the length of the prediction was 18 years. The results for the other frequencies are inconclusive as neither showed significantly better performances for either model. Including drought improved the forecasting accuracy in most instances for barley, soybeans and wheat, but noticeably diminished the model performance for cocoa. The rest of the commodities showed improvements for some frequency scales but a negative influence for others.

To conclude, the addition of the frequency domain to the analysis revealed valuable information about the cyclical behavior of both the commodity prices and the drought index, which would have otherwise been left unnoticed. Therefore, the spectral factor models are recommended to be used not only on company specific financial data as done by Bandi et al. (2019), but also for different fields of research. Regarding the forecasting, the inclusion of drought leads to better accuracy of the estimates in a majority of the testing scenarios. However a number of cases, which experienced diminished performance, showcase the shortcomings and call for extended analysis before including the variable.

Bibliography

Ababio, K. A. (2012). *Comparative Study of Stock Price Forecasting Using Arima and Arimax Models*. PhD thesis, Kwame Nkrumah University of Science and Technology, Kumasi.

Al-Sudani, H. (2019). Temperature − Potential Evapotranspiration Relationship in Iraq Using Thornthwaite Method. *Journal of University of Babylon for Engineering Sciences*, 27:16–25.

Allen, R. G., Pereira, L. S., Raes, D., Smith, M., et al. (1998). Crop evapotranspiration – Guidelines for computing crop water requirements – FAO Irrigation and drainage paper 56. *Fao, Rome*, 300(9):D05109.

Arndt, D. S. and Blunden, J. (2012). State of the Climate in 2011. *Bulletin of the American Meteorological Society*, 93(7).

Bachmair, S., Tanguy, M., Hannaford, J., and Stahl, K. (2018). How well do meteorological indicators represent agricultural and forest drought across Europe? *Environmental Research Letters*, 13(3):34042.

Baffes, J. and Haniotis, T. (2010). Placing the 2006/08 Commodity Price Boom into Perspective. *The World Bank, Policy Research Working Paper Series.*

Bandi, F. M., Chaudhuri, S., Lo, A. W., and Tamoni, A. (2019). Spectral factor models. *Johns Hopkins Carey Business School Research Paper*, (No. 18-17).

Baxter, M. and King, R. G. (1995). Measuring Business Cycles Approximate Band-Pass Filters for Economic Time Series. Working Paper 5022, National Bureau of Economic Research.

Beaudry, P., Galizia, D., and Portier, F. (2017). Is the Macroeconomy Locally Unstable and Why Should We Care? *NBER Macroeconomics Annual*, 31:479–530.

Bellemare, M. F. and Carnes, N. (2015). Why do members of congress support agricultural protection? *Food Policy*, 50:20–34.

Benhabib, J. and Nishimura, K. (1979). The hopf bifurcation and the existence and stability of closed orbits in multisector models of optimal economic growth. *Journal of Economic Theory*, 21(3):421–444.

Benhabib, J. and Nishimura, K. (1985). Competitive equilibrium cycles. *Journal of Economic Theory*, 35:284–306.

Briffa, K., Schrier, G., and Jones, P. (2009). Wet and dry summers in Europe since 1750: evidence of increasing drought. *International Journal of Climatology*, 29:1894–1905.

Brooks, M., Mattos, F., and Schoengoldtos, K. (2014). How Do Agricultural Futures Prices Respond To New Information About Drought Conditions? In *NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*, St. Louis, MO.

Burkhead, T. and Klink, V. (2018). American agricultural commodities in a changing climate. *AIMS Agriculture and Food*, 3:406–425.

Calvo, J. C. (1986). An evaluation of Thornthwaite's water balance technique in predicting stream runoff in Costa Rica. *Hydrological Sciences Journal*, 31(1):51–60.

Canova, F. (1998). Detrending and business cycle facts. *Journal of Monetary Economics*.

Carr, M. and Lockwood, G. (2011). The water relations and irrigation requirements of cocoa (Theobroma cacao L.): a review. *Experimental agriculture*, 47(4):653–676.

Cashin, P., McDermott, C. J., and Scott, A. (1999). Booms and Slumps in World Commodity Prices. *World Development*, pages 1–24.

Cogley, T. and Nason, J. M. (1995). Effects of the Hodrick-Prescott filter on trend and difference stationary time series Implications for business cycle research. *Journal of Economic Dynamics and Control*, 19(1):253–278.

Congressional Budget Office (2017). Options to Reduce the Budgetary Costs of the Federal Crop Insurance Program.

Dai, A. (2011). Drought under global warming: a review. *WIREs Climate Change*, 2(1):45–65.

Danthine, J.-P. and Girardin, M. (1989). Business cycles in Switzerland: A comparative study. *European Economic Review*, 33(1):31–50.

Davidson, R., Labys, W., and Lesourd, J.-B. (1997). Wavelet Analysis of Commodity Price Behavior. *Computational Economics*, 11:103–128.

Day, R. H. and Shafer, W. (1987). Ergodic fluctuations in deterministic economic models. *Journal of Economic Behavior and Organization*.

Deo, R. and Şahin, M. (2015). Application of the Artificial Neural Network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using hydrometeorological parameters and climate indices in eastern Australia. *Atmospheric Research*, 161-162.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical science*, pages 189–212.

Dragotă, C.-S., Dumitraşcu, M., Grigorescu, I., and Kucsicsa, G. (2011). The Climatic Water Deficit in South Oltenia Using the Thornthwaite Method. In *Forum geografic*, volume 10.

Ďurka, P. and Pastoreková, S. (2012). ARIMA vs. ARIMAX–which approach is better to analyze and forecast macroeconomic time series. In *Proceedings of the 30th International Conference Mathematical Methods in Economics, Karviná, Czech Republic*, pages 11–13.

Elder, J. and Jin, H. J. (2007). Long memory in commodity futures volatility: A wavelet perspective. *Journal of Futures Markets*, 27(5):411–437.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*.

Federal Emergency Management Agency (FEMA) (1995). National Mitigation Strategy: Partnerships for Building Safer Communities. Mitigation Directorate. Technical report, Washington, DC: Federal Emergency Management Agency.

Folland, C., Karl, T. R., and Salinger, M. (2006). Observed climate variability and change. *Weather*, 57:269–278.

Fourier, J. B. J. (1822). *Théorie analytique de la chaleur*. Paris: Firmin Didot, père et fils.

Freire-González, J., Decker, C., and Hall, J. W. (2017). The Economic Impacts of Droughts: A Framework for Analysis. *Ecological Economics*.

Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, 93(26):429–441(12).

Granger, C. and Stărică, C. (2005). Nonstationarities in Stock Returns. *The Review of Economics and Statistics*, 87:503–522.

Guttman, N. B. (1998). Comparing the Palmer Drought Index and the Standardized Precipitation Index. *JAWRA Journal of the American Water Resources Association*, 34(1):113–121.

Hamilton, J. D. (2017). Why You Should Never Use the Hodrick-Prescott Filter. Working Paper 23429, National Bureau of Economic Research.

Harvey, A. and Jaeger, A. (1993). Detrending, Stylized Facts and the Business Cycle. *Journal of Applied Econometrics*, 8(3):231–47.

Hodrick, R. and Prescott, E. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, 29(1):1–16.

Hodrick, R. J. and Prescott, E. C. (1980). *Post-war US Business Cycles: An Empirical Investigation*. Center for Mathematical Studies in Economics and Management Science, Northwestern University.

Hudson, D. T. (2019). *An analysis of the Standardized Precipitation Evapotranspiration Index (SPEI) on drought based on Economic Impacts in the Northeast (US)*. PhD thesis, City College of New York.

Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688.

ICCO (1999). Annual Report for 1998/99.

IPCC (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: Special report of the intergovernmental panel on climate change*. [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 582 pp.

IPCC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Juglar, C. (1862). *Des crises commerciales et de leur retour périodique en France: en Angleterre et aux États-Unis.* Guillaumin e cie.

Kamara, A., Korajczyk, R. A., Lou, X., and Sadka, R. (2016). Horizon Pricing. *Journal of Financial and Quantitative Analysis*, 51(6):1769 – 1793.

Kang, B. U., In, F. H., and Kim, T. S. (2012). Timescale Betas and the Cross Section of Equity Returns: Framework, Application, and Implication for Interpreting the Fama-French Factors. *SSRN Electronic Journal.*

Kaplan, I. (2001). Applying the Haar Wavelet Transform to Time Series Information.

Kitchin, J. (1923). Cycles and Trends in Economic Factors. *The Review of Economics and Statistics*, 5(1):10–16.

Kondratiev, N. D. (1925). The major economic cycles. *Voprosy Konjunktury*, 1(1):28–79.

Kongcharoen, C. and Kruangpradit, T. (2013). Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) Model for Thailand Export. In *the 33rd International Symposium on Forecasting*, Seoul.

Koopman, S. J. and Lee, K. M. (2005). Measuring Asymmetric Stochastic Cycle Components. *Available at SSRN 785024.*

Kraemer, R. A. (2007). *Economic Impact of Droughts: Challenges for Water & Environmental Policies.* Water Scarcity and Drought - A Priority of the Portuguese Presidency.

Kriechbaumer, T., Angus, A., Parsons, D., and Casado, M. R. (2014). An improved wavelet–ARIMA approach for forecasting metal prices. *Resources Policy*, 39:32–41.

Kupkanchanakul, T. (2000). Expert Consultation on Bridging the Rice Yield Gap in the Asia-Pacific Region. Technical report, Food and Agriculture Organization.

Kuwayama, Y. (2019). The Economic Impacts of Drought on US Agriculture.

Kuznets, S. (1930). *Secular movements in production and prices; their nature and their bearing upon cyclical fluctuations*. Houghton Mifflin, Boston.

Kydland, F. and Prescott, E. (1990). Business cycles: real facts and a monetary myth. *Quarterly Review*, 14(Spr):3–18.

Labys, W. C., Kouassi, E., and Terraza, M. (2000). Short-term cycles in primary commodity prices. *Developing Economies*, 38(3):330–342.

Le, V. P. M., Meenagh, D., and Minford, P. (2019). A long-commodity-cycle model of the world economy over a century and a half — Making bricks with little straw. *Energy Economics*, 81:503–518.

Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859.

Ledoit, O. and Wolf, M. (2011). Robust performances hypothesis testing with the variance. *Wilmott*, 2011(55):86–89.

Light, J. and Shevlin, T. (1998). The 1996 grain price shock: how did it affect food inflation. *Monthly Lab. Rev.*, 121:3.

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, 47(1):13–37.

Long, J. S. and Ervin, L. H. (2000). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *The American Statistician*, 54(3):217–224.

Mallikarjuna, M. and Rao, R. P. (2019). Evaluation of forecasting methods from selected stock market returns. *Financial Innovation*, 5(1):40.

McKee, T. B., Doesken, N. J., Kleist, J., et al. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, pages 179–183. Boston.

Mitchell, J. F. B. and Karoly, D. J. (2001). Detection of Climate Change and Attribution of Causes. *The Third IPCC Assessment Report. Working Group I: The Scientific Basis*.

Mitchell, W. C. (1927). *Business Cycles: The Problem and Its Setting*. NBER.

Modirroosta, M. (2013). A decomposition analysis of base metal prices: Comparing the effect of detrending methods, on trend identification and cyclical components. Master's thesis, School of Business and Economics.

National Centers for Environmental Information (NCEI) (2020). Calculating the Cost of Weather and Climate Disasters.

Neukom, R., Steiger, N., Gómez-Navarro, J. J., Wang, J., and Werner, J. P. (2019). No evidence for globally coherent warm and cold periods over the preindustrial Common Era. *Nature*, 571(7766):550–554.

Otkin, J., Svoboda, M., Hunt, E., Ford, T., Anderson, M., Hain, C., and Basara, J. (2017). Flash Droughts: A Review and Assessment of the Challenges Imposed by Rapid-Onset Droughts in the United States. *Bulletin of the American Meteorological Society*, 99.

Palmer, W. C. (1965). *Meteorological drought*, volume 30. US Department of Commerce, Weather Bureau.

Panchal, A. and Patel, J. (2017). The Success Prediction Ratio (SPR): A Simplified Generalized Technique to Compare and Validate the Performance of the Stock Price Prediction Model. *International Journal for Research in Applied Science and Engineering Technology*, 5:797–805.

Park, G. (1996). The role of detrending methods in a model of real business cycles. *Journal of Macroeconomics*, 18(3):479–501.

Patton, A., Politis, D. N., and White, H. (2009). Correction to "Automatic Block-Length Selection for the Dependent Bootstrap" by D. Politis and H. White. *Econometric Reviews*, 28(4):372–375.

Pedersen, T. M. (2003). Alternative Linear and Non-Linear Detrending Techniques: A Comparative Analysis based on Euro-Zone Data. *Ministry of Economic and Business Affairs*, pages 1–46.

Piesse, J. and Thirtle, C. (2009). Three bubbles and a panic: An explanatory review of recent food commodity price events. *Food Policy*.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical association*, 89(428):1303–1313.

Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70.

Potopová, V., Boroneant, C., Možný, M., Štěpánek, P., and Skalák, P. (2013). Observed spatiotemporal characteristics of drought on various time scales over the Czech Republic. *Theoretical and Applied Climatology*, 115:563–581.

Potopová, V., Štěpánek, P., Možný, M., Türkott, L., and Soukup, J. (2015). Performance of the standardised precipitation evapotranspiration index at various lags for agricultural drought risk assessment in the Czech Republic. *Agricultural and Forest Meteorology*, 202:26–38.

Prabnakorn, S., Maskey, S., Suryadi, F., and Fraiture, C. (2018). Rice yield in response to climate trends and drought index in the Mun River Basin, Thailand. *Science of The Total Environment*, 621:108–119.

Pradesh, A., Venkataramanaiah, M., and Campus, G. V. I. (2018). Forecasting time series stock returns using ARIMA: Evidence from S&P BSE SENSEX. *International Journal of Pure and Applied Mathematics*, 118(24).

Ramsey, J. (1999). The contribution of wavelets to the analysis of economic and financial data. *Philosophical Transactions of The Royal Society B: Biological Sciences*, 357.

Riebsame, W., Changnon, S., and Karl, T. (1991). *Drought and Natural Resources Management in the United States: Impacts and Implications of the 1987-89 Drought*. Westview Press.

Riedl, B. (2002). Agriculture Lobby Wins Big in New Farm Bill. Technical report, The Heritage Foundation.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*.

Schüler, Y. (2019). On the Cyclical Properties of Hamilton's Regression Filter and Refinements. *SSRN Electronic Journal*.

Schwartz, E. and Smith, J. E. (2000). Short-term variations and long-term dynamics in commodity prices. *Management Science*, 46(7):893–911.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, 19(3):425–442.

Singleton, K. J. (1988). Econometric issues in the analysis of equilibrium business cycle models. *Journal of Monetary Economics*, 21(2-3):361–386.

Solomon, S., Daniel, J. S., Sanford, T. J., Murphy, D. M., Plattner, G.-K., Knutti, R., and Friedlingstein, P. (2010). Persistence of climate changes due to a range of greenhouse gases. *Proceedings of the National Academy of Sciences*, 107(43):18354–18359.

Törnros, T. and Menzel, L. (2014). Addressing drought conditions under current and future climates in the Jordan River region. *Hydrology and Earth System Sciences*, 18(1):305–318.

Trenberth, K. E., Dai, A., Rasmussen, R. M., and Parsons, D. B. (2003). The Changing Character of Precipitation. *Bulletin of the American Meteorological Society*, 84(9):1205–1218.

Trostle, R. (2011). *Why Have Food Commodity Prices Risen Again?* DIANE Publishing Company.

United States Department of Agriculture (1997). *Usual planting and harvesting dates for U.S. field crops* . U.S. Dept. of Agriculture, National Agricultural Statistics Service Washington, D.C.

United States Department of Agriculture (2019). Irrigation & Water Use. Technical report, USDA.

Vagropoulos, S., Chouliaras, G., Kardakos, E., Simoglou, C., and Bakirtzis, A. (2016). Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting. In *2016 IEEE International Energy Conference (ENERGYCON)*, pages 1–6.

Varouchakis, E. and Corzo, G. (2019). *Spatiotemporal Analysis of Extreme Hydrological Events*. Elsevier.

Vavřina, M. (2012). *Charles University in Prague Faculty of Social Sciences Institute of Economic Studies Volatility Spillovers in New Member States : A Bayesian Model*. PhD thesis, Charles University.

Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, 23(7):1696–1718.

Wells, N., Goddard, S., and Hayes, M. (2004). A Self-Calibrating Palmer Drought Severity Index. *Journal of Climate - J CLIMATE*, 17:2335–2351.

Wessel, M. and Quist-Wessel, P. M. F. (2015). Cocoa production in West Africa, a review and analysis of recent developments. *NJAS - Wageningen Journal of Life Sciences*, 74-75:1–7.

Whittaker, E. T. (1923). On a new method of graduation. In *Proceedings of the Edinburgh Mathematical Society*, volume 41, pages 63–75.

Wilhite, D. A. and Glantz, M. H. (1985). Understanding: the Drought Phenomenon: the Role of Definitions. *Water international*, 10(3):111–120.

Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Almqvist and Wiksell, Sweden.

World Bank (2020). World Bank Commodities Price Data (The Pink Sheet). `http://www.worldbank.org/commodities`. Accessed on 10.02.2020.

World Meteorological Organization (2012). *Standardized Precipitation Index User Guide*. WMO.

World Meteorological Organization (2014). WMO statement on the status of the global climate in 2013. Technical Report No. 1130, WMO.

Wu, Z., Huang, N. E., Long, S. R., and Peng, C. K. (2007). On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences of the United States of America*, 104(38):14889–14894.

Zhou, T. and Hong, T. (2013). Projected Changes of Palmer Drought Severity Index under an RCP8.5 Scenario. *Atmospheric and Oceanic Science Letters*, 6(5):273–278.

Zipper, S. C., Qiu, J., and Kucharik, C. J. (2016). Drought effects on US maize and soybean production: spatiotemporal patterns and historical changes. *Environmental Research Letters*, 11(9):94021.