POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Modelling Spread in European Union Allowances Market

Tesi di Laurea Magistrale in
Mathemetical Engineering - Ingegneria Matematica

Author: **Christopher Testa**

Student ID: 976744
Advisor: Prof. Roberto Baviera
Co-advisors: Pietro Manzoni
Academic Year: 2022-23

Abstract

This thesis focuses on the analysis of the spread in the European Union Allowances futures market, these are contracts created at European level to regulate greenhouse gas emissions from power plants and industrial factories. In particular, one contract allows compliance entities to compensate the emission of one ton of CO_2.

As observed in different studies, a spread term is applied on the risk-free rate of carbon futures contracts. The proposed study aims to find the main drivers of this spread, since it can cause arbitrage opportunities.

To study this process, the time window has been split in two parts, the so-called Phase 2 (2008-2012) and Phase 3 (2013-2020) as the shift from one phase to the other came with important changes in the European Emission Trading System framework. The analysis proposes different regression models applied to the spread values. The variables used as regressors of the model are economic factors, commodities-related variables, and the Z-spread of compliance companies.

This analysis is proposed under two lights. On one side all the available observations are considered, while on the other, only the ones with volume over a certain threshold are selected, to focus on the liquid market.

Our results show that using filtered datasets instead of complete datasets can lead to different outcomes when applying models. However, we can confirm that the main drivers of the spread, especially in Phase 3, are the short-term rate, the volatility of spot returns, and the Z-spread. The last analysis proposed studies the spread as an autoregressive (AR, hereinafter) phenomenon. We can confirm that it behaves like an AR(1) event, which tends to be negative for the majority of its lifespan and converges to zero as contract maturity approaches.

Keywords: EU-ETS, spread, carbon futures, Z-spread, autoregressive model

Abstract in lingua italiana

Questo lavoro di tesi si concentra sull'analisi dello spread rilevato nel mercato dei futures relativi ai contratti European Union Allowances, ovvero contratti creati a livello Europeo per regolare le emissioni di gas serra di centrali elettriche e impianti industriali. In particolare, un contratto consente alle aziende soggette a regolazione di compensare l'emissione di una tonnellata di CO_2.
Come osservato in diverse ricerche, esiste uno spread applicato al tasso privo di rischio nei contratti carbon futures. L'analisi presentata mira a ricavare i principali fattori che guidano questo spread, dato che quest'ultimo può dare origine ad opportunità di arbitraggio.
Per studiare il processo, la finestra temporale è stata divisa in due parti, definite Fase 2 (2008-2012) e Fase 3 (2013-2020) dato che il passaggio da una fase all'altra è stato accompagnato da importanti variazioni nel contesto legislativo dell'European Emission Trading System. L'analisi propone diversi modelli di regressione, applicati sui valori di spread. Le variabili usate come regressori del modello sono fattori economici, variabili legate alle commodities e lo Z-spread delle aziende regolamentate.
L'analisi viene proposta sotto due diverse luci. Da una parte tutte le osservazioni disponibili sono considerate, mentre dall'altra, solo quelle che hanno volume sopra a una certa soglia sono considerate, per concentraci sul mercato liquido.
I nostri risultati dimostrano che l'uso di dataset filtrati anziché completi può produrre risultati diversi quando si applicano i modelli. Tuttavia, possiamo confermare che i principali fattori che guidano lo spread, soprattutto in Fase 3, sono il tasso a breve termine, la volatilità dei rendimenti spot e lo Z-spread. L'ultima analisi proposta studia lo spread come un fenomeno autoregressivo (AR). Possiamo confermare che esso si comporta come un evento AR(1), il quale tende a rimanere stabilmente negativo per la maggior parte del tempo, per poi convergere a zero quando ci si avvicina alla scadenza del contratto.

Parole chiave: EU-ETS, spread, carbon futures, Z-spread, modello autoregressivo

Notation and Symbols

$F_{t_0,T}$	Futures price at time t_0
S_{t_0}	Spot price at time t_0
$R_{t_0,T}$	Interest rate of the dealer at time t_0
$r_{t_0,T}$	Risk-free rate between t_0 and T
$s_{t_0,T}$	Spread rate between t_0 and T
$CY_{t_0,T}$	Convenience yield between t_0 and T
$R^{OIS}(t_0, t_e)$	OIS Rate at time t_0
$B(t_0, t_e)$	Risk-free discount factor at t_0 for time t_e
r_i	EUA Spot prices return at time i
σ_t	Standard deviation of spot returns at time t
$Z(t_0, T)$	Z-spread at time t_0 for maturity T
$\overline{B}(t_0, t_e)$	Corporate bonds discount factor
$C_{t_0,T}$	Coupon bond clean price
Acc	Accrual term
c	Coupon value
δ_i	Time interval between t_{i-1} and t_i
G	Synthetic forward
K	Strike price

Acronyms

AR	Autoregressive
GHG	Greenhouse Gas
EUAs	European Union Allowances
CY	Convenience Yield
EU-ETS	European Emission Trading System
EEX	European Energy Exchange
ICE	Intercontinental Exchange
OHA	Operator Holding Account
NAPs	National Allocation Plans
CER	Certified Emission Reduction
ERU	Emission Reduction Unit
ECB	European Central Bank
EUTL	European Union Transaction Log
OIS	Overnight Index Swap
OLS	Ordinary Least Squares
ACF	Autocorrelation Function
PACF	Partial Autocorrelation Function
ADF	Augmented Dickey–Fuller

Contents

Introduction

The European Union, worried about the global warming, is trying to reduce the emissions of Greenhouse Gases (GHGs, hereinafter) caused by stationary installations. The first step in this direction was taken in 2000, when the European commission discussed the implementation of a cap-and-trade system aimed at limiting gas emissions as well as encouraging the EU members to respect their Kyoto protocol obligations (Ellerman et al. (2016)). Almost five years later, the European Union Allowances (EUAs) market was established.

This market operates through contracts known as allowances. As the name suggests, they allow the owner to emit one ton of CO_2 equivalent of GHGs. The market operates on the principle that if a factory emits GHGs, it must compensate for those emissions each year purchasing a proportional number of carbon credits. These contracts can be obtained in different ways. One way is directly from the European Union which, through national plans (previously) and with the use of benchmarks (currently), provides companies with free allowances. Another way is through auctions, regularly held for certified market participants. These two methods represent the primary market, while spot contracts, futures and options on futures represent the secondary market (ESMA (2022)).

After an initial pilot phase between 2005 and 2007, during which the free allocations exceeded the number of contracts needed to compensate emissions, to make the market work, the EU decided to reduce the amount of contracts given for free, and the market became effective. Furthermore, since 2013, the EU implemented major changes that made the market even more effective. These changes include the reduction of free allocations for power producers to zero, as they represent the sector with the most emissions (Ellerman et al. (2016)), and the introduction of a fixed annual rate of reduction of free allocations.

When analysing the secondary market, one can see two main relevant vehicles to obtain the contracts, the spot market and the futures market.

An interesting fact is that the relationship between the two prices reveals the presence of a spread term over the risk-free rate. The analysis of this spread is important because its existence can create arbitrage opportunities.

The purpose of this thesis is to analyse the main factors driving this phenomenon.

Multiple researches have followed the evolution of the spread between carbon spot price and futures price (Palao and Pardo (2021); Trück and Weron (2016)) defining the said spread as a convenience yield (CY). This study builds upon the work of Trück and Weron (2016) by expanding their analysis over time and incorporating additional variables into their models, with the objective of gathering more information about the phenomenon.

The thesis aims to:

- Replicate and extend the models presented in Trück and Weron (2016) to a different time window, including Phase 3, and analyze the results obtained by solely considering observations with high volumes to select liquid contracts.
- Analyse the relationship between commodity related variables and the spread in EUAs market. As literature (Alberola et al. (2008)) focused purely on analyzing the relation between commodity variables and the spot price of allowances.
- Study the relationship between the spread in EUAs market and the Z-spread, which represents the spread between the risk-free rate and the rate of return of bonds issued by compliance companies.
- Verify whether the spread can be considered as an autoregressive process, and in case, find the variables that can contribute to the explanation of the process.

The study is organised into two main sections, the first one is focused on the analysis of the time window 2008-2012, which represents Phase 2, and the second one is focused on the period 2013-2020, that defines Phase 3.

The rest of the thesis is structured as follows:

Chapter 1 provides a general overview of the market, describing it from a financial perspective and tracing its historical evolution, while also offering insights on the relevant literature considered.

Chapter 2 gives the details on the dataset used, providing the data sources and the relevant information related to them.

Chapter 3 defines the methodology used, giving the reader the main formulas used and defining the models that have been implemented.

Chapter 4 is focused on replicating the analyses proposed in the literature and extending previous models. This is achieved by introducing new variables and shifting the focus to observations derived from liquid transactions.

Chapter 5 focuses on applying the presented models to Phase 3 and introducing other variables for both liquid and non-liquid cases. The chapter also proposes a study related to the Z-spread of compliance entities and an autoregressive model relative to a rolled version of the spread. Additionally, the chapter attempts to analyze the options market.

Chapter 6 summarizes the work and provides conclusive comments.

1 | Financial Framework

In this first chapter the market of European Union Allowances is described, defining the contracts traded and their characteristics in detail. Moreover, it is presented the historical evolution of the market and the two main exchanges on which these contracts are traded. Finally, it is given an overview of the literature used for this analysis.

1.1. Contracts Description

European Union Allowances are contracts created to reduce the emission of greenhouse gases at European level. These pseudo-commodities allow the owner to compensate the emission of 1 ton of CO_2 per contract. This market is relatively new, and after a first unsuccessful period (2005-2007) it has set the basis for GHG emissions reduction in Europe. The European Union Emission Trading System (EU-ETS) was established in 2005 and quickly became the world's leading market for carbon trading. The EU-ETS is designed as a cap-and-trade system, meaning that a limited number of contracts are issued every year (cap) and that companies can trade them with other compliance entities, or buy additional contracts via auctions or via the secondary market (trade).
The main market participants are compliance entities, which are companies that face the obligation to close a certain number of contracts to compensate their yearly GHGs emissions. In fact, every year at the end of December, each compliance company must submit their "observed emissions", a data that represents the emissions monitored by the company during the year. These will be verified at national level by some EU delegated regulators by the end of the following year's March. Companies have then time till the end of April to close as many contracts as needed to match the "verified emissions". If this obligation is not met, the company will be subject to a penalty fee proportional to the amount of CO_2 equivalent not compensated.
Compliance entities can obtain EUAs in three different ways: via free allocations, by auctions or operating in the secondary market. Regarding the first methodology, as later specified, companies receive free allowances according to what the European Council considers appropriate, following determined benchmarks and corrections.

For what concerns auctioning, from 2012, the European Energy Exchange (EEX) has been delegated by the European authorities as main platform for EUAs auctions. In fact, every Monday, Tuesday and Thursday, selected market participants can take part in the auctions of allowances. As of 2013 this market became the dominant way for companies to obtain EUAs.
Finally, the last vehicle to get these contracts is via futures trading.
EUAs futures contracts have physical delivery and are traded in lots. Each lot represents 1000 EUAs, meaning one future allows for the emission of 1000 tons of CO_2. Focusing on the economic part, the price is quoted in euro cents per metric ton, with minimum tick of 0.01 euro[1], which decreased from previous minimum tick of 0.05 euro in March 2007 (Frino et al. (2010); Pro (2022)).
The most active futures exchange is represented by the Intercontinental Exchange (ICE) (Palao and Pardo (2021)), and the most traded futures contract every year is the one maturing in December. For what concerns the settlement price procedure in ICE, these contracts have settlement period in the time span between 16:50:00 and 16:59:59 UK local time, with the price being a weighted average of the trading prices when the number of lots traded in the closing period exceeds 100. In alternative, the price is determined by a Market Supervision Official as either an average of quoted prices provided by market participants or as a price determined by the Market Supervision Official considering previous prices on ICE Platform trading session[2]. The price can be adjusted before the publishing, as to avoid arbitrage.
Once companies successfully obtain allowances contracts, to use them they have to open an Operator Holding Account (OHA). From 2012 the European Union centralised all operations to a single registry: The Union Registry, which holds data relative to all the countries participating in the EU-ETS[3]. To open an OHA a compliance entity has to pay a fee, which is negligible compared to big firms' values, and nominate two representatives when the account is first activated. Via the Union Registry accounts, aircraft[4] and stationary operators can do three fundamental things: receive the allowances they have allocated or trade them; submit and approve the verified CO_2 emission on the European Registry; and finally surrender allowances to compensate the verified emissions.
The deadlines compliance entities must respect are:

---

[1]More information regarding ICE Futures contracts can be found in the *ICE EUA Futures Contract Specifications: https://www.theice.com/products/197/EUA-Futures* (accessed on 2 December 2022).

[2]Further information is available in the ICE ECX Contracts: EUAs and CERs User Guide, ECX (2010).

[3]For more information consult the *Union Registry: https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets/union-registry_ en* (accessed on 10 March 2023).

[4]Aircraft operators have a designed market of European Union Aviation Allowances EUAAs

- January 1, compile the Annual Emission Report, which records the emissions of the previous year that will be verified by authorized authorities.
- Not later than March 31, submit the previous year's verified emissions.
- Not later than 30 April, surrender the needed allowances.

1.2. Phases in the EU Emission Trading System

The EU-ETS is divided in different phases, that highlight specific time windows. From 2005 to 2007 the market opened in the so-called Phase 1, which represents an initial experiment of the cap-and-trade system. This phase is considered a failure, as basically all free allocations were enough to support the emissions of companies, consequently the spot price of EUAs fell to zero at certain times (Swinkels and Yang (2022)). These free allocations were calculated through "National Allocation Plans" (NAPs) in which each country had to submit their predicted emissions to the European Commission who evaluated the plan and decided if it satisfied the guiding documents[5] (Clarkson et al. (2015)).

A key feature of this new market is that compliance entities which could not submit the allowances in time had a penalty fee of 40 euros per ton of CO_2 not compensated.

In the end Phase 1 did not achieve any particular result, but it succeeded in establishing a carbon price and setting the basis of this newly introduced market.

Going forward in time we reach Phase 2, which describes the time interval from the beginning of 2008 to the end of 2012 and represents a turning point compared to the previous period. The reason for this is that allowances from Phase 1 could not be used to compensate emissions in this phase, so the market had a sort of restart.

The second change that influenced the market is the reduction of the freely allocated allowances, which diminished to almost 90% of the previous phase allocations. To the market were also added three new countries: Iceland, Liechtenstein, and Norway.

The European Commission tried to make NAPs more transparent and simpler[6], and a lot of proposed plans saw their predicted allocations reduced. However, again, most of the needed allowances were distributed for free.

In this phase, allowances began to be auctioned. This trading method got more and more popular with time, becoming the main market vehicle from 2013. From compliance

[5] Phase 1 guidance can be found in *Communication from the Commission: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52003DC0830&from=EN* (accessed on 28 March 2023).

[6] Phase 2 guidelines in *Further guidance on allocation plans for the 2008 to 2012 trading period: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52005DC0703&from=EN* (accessed on 28 March 2023).

entities' point of view an important change introduced was the increase of the penalty fee to 100 euros per ton of CO_2 not compensated. A new key feature was also the possibility to bank the exceeding allowances, so that contracts that were not submitted in a certain year could be carried on in time to compensate future emissions.
The following phase, Phase 3, lasted from 2013 to December 2020. Different changes were made in this period, one of most relevant is the dismissal of the NAPs in favor of the use of benchmarks. These are calculated at European level, based on the 10% best performing installations in the market sector[7].
From trading perspective auctioning became the default method to allocate allowances, since the number of free allocations drastically diminished, as can be seen from Figure 1.1.

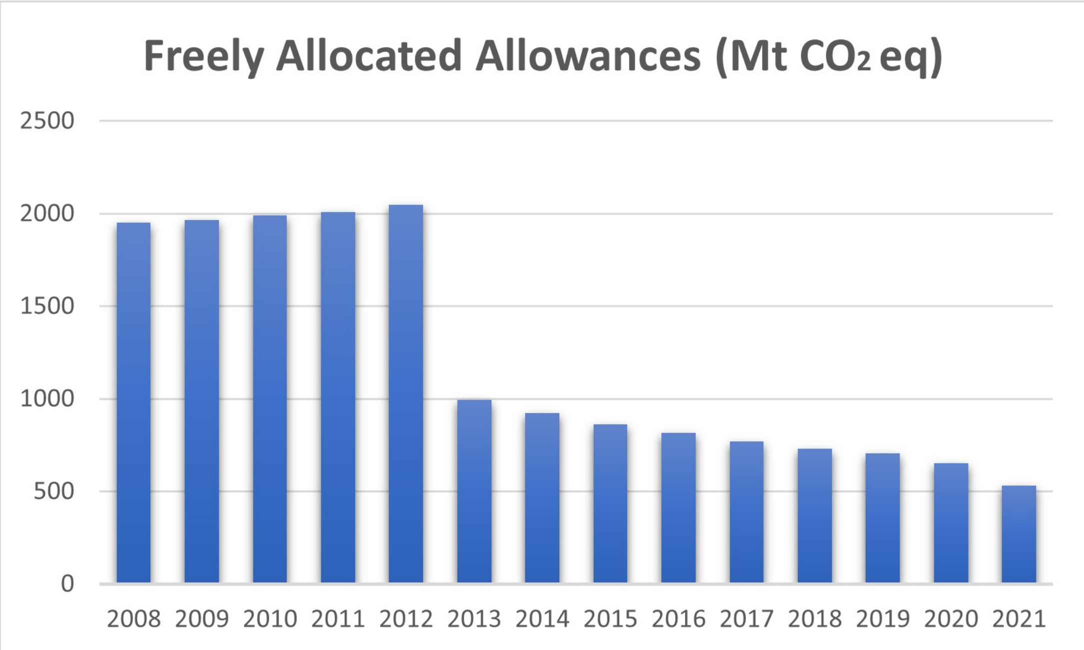
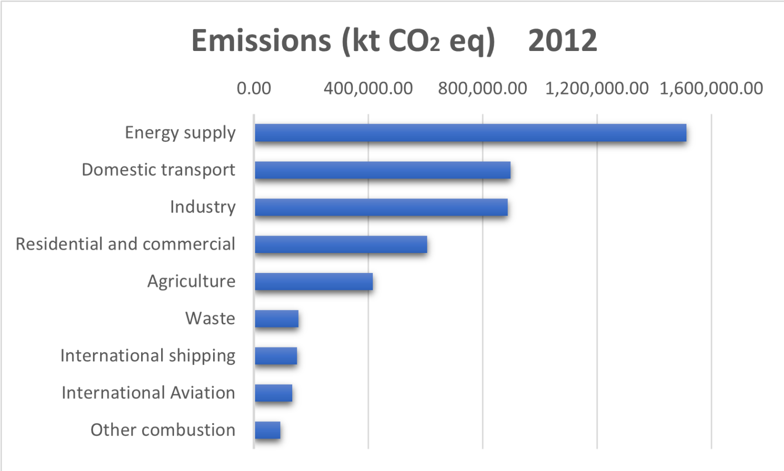


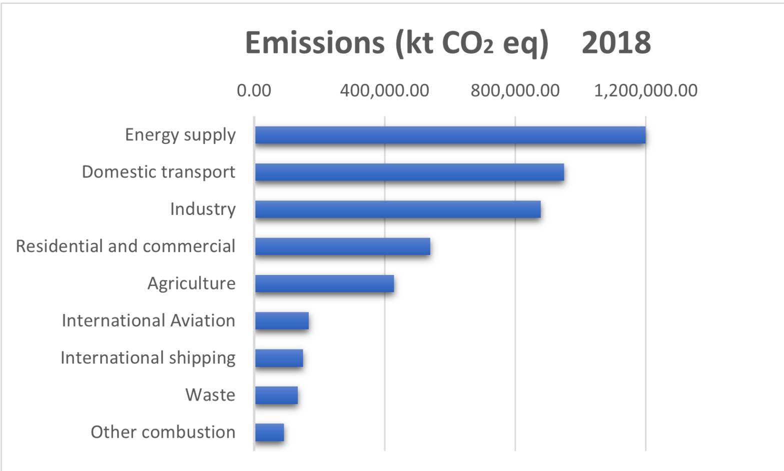
Figure 1.1: Graph of free allocations (in M-tons of CO_2 equivalent) per year at European level. Source: European Environment Agency[8].

The reason for this drop is that free allowances were no longer distributed to power generators, moreover, manufacturing industry received each year a lower portion of free allowances. Power generators were selected for this cut as they represent the sector with most GHG produced, confirmation in this sense derive from Figure 1.2.

[7]More information about this can be found on *Official Journal of the European Union: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32011D0278* (accessed on 10 March 2023).
[8]The European Environment Agency provides a data viewer of the emission trading data present in the European Union Transaction Log. Its website is *https://www.eea.europa.eu/data-and-maps/dashboards/emissions-trading-viewer-1* .



(a) Emissions by sector in 2012



(b) Emissions by sector in 2018

Figure 1.2: Graphs of emissions (in kilotons of CO_2 equivalent), divided by sector, for 2012 and 2018. Source for both figures: European Environment Agency[9].

Finally, a decline rate of 1.74% was set at the beginning of the phase and applied to all free allocated allowances. This means that the amount of freely given contracts linearly declines each year.
Another relevant feature introduced in Phase 3 is that the acceptability of offset contracts was limited, meaning Certified Emissions Reductions (CERs) and Emission Reduction Units (ERUs) contracts were less and less accepted, contributing to centralize the EU-ETS market on allowances.
CERs are international credits produced through the Clean Development Mechanism, one

[9]The European Environment Agency provides a data viewer of the GHG emissions sent to the United Nations and the the EU Greenhouse Gas Monitoring Mechanism. Its website is *https://www.eea.europa.eu/data-and-maps/data/data-viewers/greenhouse-gases-viewer* .

of the mechanisms planned by the Kyoto protocol to reduce GHG emissions. To receive these credits some steps have to be followed, a public or private agent has to design a project to limit the gas emissions and derive the difference between the real emitted gas quantity and the emissions realised without the implementation of the plan. This difference is then considered as avoided emissions and stored as CERs[10].

ERUs are obtained in a similar way, they are the direct consequence of the Joint Implementation mechanisms. In these, two nations are involved, in particular, a public or private agent from a nation designs a project for the other country, aimed at the reduction of GHGs emissions. As before, the difference between the expected emissions before the implementation of the plan and the realised emissions gets stored as ERUs[11].

The period we are currently in is the so-called Phase 4. The shift in phase saw fewer changes compared to the previous one, in fact the only major variation is the increase in the free allocations diminishing yearly rate from 1.74% to 2.2%. This change was made to meet the Paris agreement condition to reduce the increase of the global temperature from 2 degrees Celsius to 1.5 above pre-industrial level and to match the 2030 target to reduce GHG emissions in Europe to 40% below 1990 level (Zaklan et al. (2021)). Lastly, some more technical aspects were revised; the main regarding rules related to carbon leakage, which represents the transfer of the production to another country with more relaxed emission constraints.

It is worth noting that there was a significant increase in spot price during the final years of Phase 3 and the early stages of Phase 4. This could potentially be attributed to the emergence of new market participants (Quemin and Pahle (2023)).

1.3. Two Exchanges

For what concerns auctioning, the European Energy Exchange was elected as the common auctioning platform by the European Commission, with current common interface called "CAP3".

Allowances contracts are auctioned on it every Monday, Tuesday and Thursday at 11 am CE(S)T, while in the intermediate days, German (weekly on Wednesdays) and Poland (bi-weekly on Fridays) allowances are auctioned, the reason behind these split stands in the opt-out of the trading system from the two nations. Germany in fact has its own trading scheme, while Poland keeps operating under EUA scheme but has its own auctioning market.

[10]For more information visit: *https://www.mase.gov.it/pagina/i-progetti-clean-development-mechanism* (accessed on 28 March 2023).

[11]For more information visit: *https://www.mase.gov.it/pagina/joint-implementation* (accessed on 28 March 2023).

EEX's headquarter is in Leipzig and represents the central point of the three markets. Each auction has a pre-established number of allowances to be sold, this also defines the market as a cap-and-trade, this quantity is determined at European level a priori every year. Not everyone can participate in this market, as EU Auctioning Regulation defines the criteria to be fulfilled by the market participants to enter the auctions[12]. In short, dealers have to be authorised at European level.
Despite being the elected platform for auctioning, EEX does not represent the main exchange for the secondary market. In fact, the Intercontinental Exchange, with headquarter in London and exchange in Amsterdam, represents the principal and most active trading platform for what concerns futures market, as most of the volume of EUA futures is traded there. Confirmations on this fact derive both from literature (Palao and Pardo (2021)), reports (ESMA (2022)) and futures contracts volume analysis, that can be seen in Figure 1.3. As a consequence, for this study ICE market has been analysed, since prices

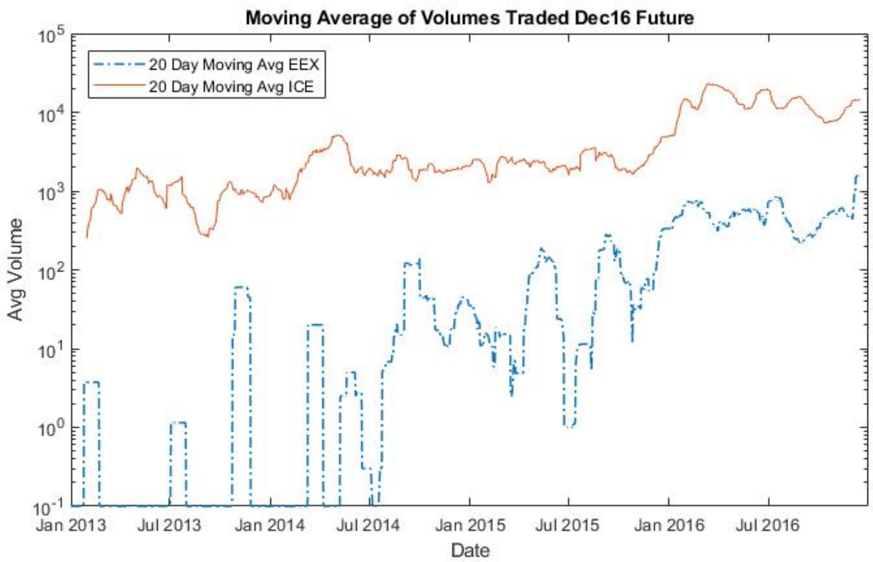


Figure 1.3: Graph of the average volume (in contract lots) traded in the last 20 days for the December 2016 future in ICE and EEX markets. When the data relative to the average volume is zero, it is set to 10^{-1}. Data Source: Refinitiv Eikon.

deriving from EEX futures are considered less reliable, because less liquid.
In this market the trading period of EUA futures contracts is between 07:00 and 17:00 UK local time as shown in ECX (2010), with a small pre-trading period lasting fifteen minutes prior to the opening hour.

[12]Full information available at *Commission Regulation: https://eur-lex.europa.eu/legal-content/EN /TXT/HTML/?uri=CELEX:02010R1031-20191128&from=EN\#tocId24* (accessed on 28 March 2023).

For what concerns futures pricing, one can take a step back and focus on the pricing of a forward contract.
We can, for example, build a static hedging portfolio composed in this way: a dealer can borrow S_0 at time t_0 and pay it back at time T, buy one lot of EUAs at price S_0 and enter into a forward contract to sell the same asset at time T. We then have the cash-flows in t_0 equal to S_0, and the ones in T as in Figure 1.4.

Figure 1.4: Cash flows of a portfolio created by borrowing S_0, purchasing one unit of asset, and selling a forward contract on the same asset.

Exploiting the no arbitrage condition, the cash flows in all time intervals must be equal, hence it is clear that the price of a forward is given by:

$$F_{t_0,T} = S_0 e^{R_{t_0,T}(T-t_0)}. \tag{1.1}$$

where $F_{t_0,T}$ is the price in t_0 for a forward contract with maturity date in T, S_0 the spot price, $e^{R_{t_0,T}(T-t_0)}$ is the actualization factor[13], and $R_{t_0,T}$ the interest rate of the dealer referred to the time interval considered.

It is worth noting that the interest rate for a typical dealer in the EUAs market is the rate at which the dealer can obtain financing, that in general is not the risk-free rate $r_{t_0,T}$. As demonstrated by market analysis, energy suppliers and transport utility are the key players in this market. These corporations do not have access to risk-free financing instruments (like the ECB's Repo auctions for large commercial banks). Instead, they rely on loans with a non-negligible credit spread. Hence, equation (1.2):

$$R_{t_0,T} = r_{t_0,T} + s_{t_0,T}. \tag{1.2}$$

[13]The year fraction of the time interval $(T - t_0)$ is used in computation, using convention Act/365.

Where $s_{t_0,T}$ represents the average spread of dealers operating within the market.
If the stochastic interest rate and the underlying certificate are statistically independent, then the futures-price process and the forward-price process $F_{t_0,T}$ in equation (1.1) are identical (Duffie (2010)).

In commodity markets, the spread term reflects the benefit of holding a good or the costs for commodity storage of the underlying commodity. Therefore, the formula used when evaluating the price of the contract can be re-written as equation (1.3):

$$F_{t_0,T} = S_0 e^{(r_{t_0,T}+s_{t_0,T})(T-t_0)}. \tag{1.3}$$

Differently from commodities, allowances do not have any cost of storage and do not pay any dividends.
It is important to underline that the spread term is not associated with any arbitrage opportunity. Instead, it refers to the credit spread that a dealer in this market has to pay pay to get financed. This spread term is evaluated in negative sign, to be coherent with the analysis made by Trück and Weron (2016) and Palao and Pardo (2021).

We aim to study how this spread behaves and which are its main drivers.

1.4. Literature Review

Different studies in this field have been proposed. In the literature, allowances are treated as a generic commodity and futures are considered as if the interest rate is the risk-free rate. This implies that its pricing follows equation (1.4):

$$F_{t_0,T} = S_0 e^{(r_{t_0,T}-CY_{t_0,T})(T-t_0)}. \tag{1.4}$$

In this, $CY_{t_0,T}$ defines the convenience yield, which represents either the advantages of holding the underlying asset, when positive, or the storage costs of the commodity, when negative (see e.g. Trück and Weron (2016); Palao and Pardo (2021)).
As previously stated in the introduction, allowances do not offer any benefits or drawbacks and come with minimal storage costs. However, the spread observed in futures markets can be interpreted as a convenience yield.

1.4.1. Convenience Yield

In the paper of Trück and Weron (2016), the authors present a general description of the EU-ETS and focus on the study of the convenience yield. Their work focuses on the analysis of Phase 2, hence considering observations of futures prices and spot prices in the time interval from April 8, 2008, to the end of 2012, selecting the futures contracts with maturity in December, as they are the most liquid, with maturity in Phase 2 and Phase 3.

They evidence that the behaviour of the market changes from backwardation[14] to contango[15] in the early years of Phase 2, after this analysis, the authors propose several regression models to study the relationship between the convenience yield and other economical regressors. The importance of a contango scenario in the commodities market lies in the fact that a robust contango, i.e. when the contango surpasses the cost of storage and interest, induces market players to buy spot and sell futures to earn from the arbitrage (Tilton et al. (2011)).

The three mainly proposed variables are the short-term interest rate, the amount of stored contracts and the volatility of spot returns.

As short-term rate the 3-month ECB quotes for AAA rated bonds is selected, the choice is done as they are considered as a proxy for the risk-free spot rate. In fact, when the authors obtained the values of spread, or convenience yield, they used ECB AAA-rated bond rates as risk-free rates. The second relevant variable chosen is the banking values, which are obtained annually by calculating the difference between the free allocations and the total verified emissions. Once this difference is calculated, the value of the previous year's banking and the offset value for the current year are added to it. This offset value is represented by the amount of CER and ERU contracts surrendered. After having obtained a yearly information, to get daily observations, a simple linear interpolation is implemented. The final variable considered is the volatility in spot returns, which is obtained via a GARCH(1,1) or via an Exponential Moving Average process to get daily observations. These variables are said to be the most relevant because the authors proposed three main hypotheses, one for each term.

The first hypothesis, on the short-term rate, defines that lower rates decrease the value of CY, suggesting a positive relationship between the two terms. This relation is deduced from the computation of the convenience yield, in fact the risk-free rate is a direct input in the CY formula, obtainable from equation (1.4). The second hypothesis, on the level of banked contracts, affirms that the benefit of holding contracts decreases as the level of storage increases, thus suggests a negative relationship between the variables. The final hypothesis presented examines the relationship between spot returns volatility and the spread, since an increase in variance would increase the demand for hedging and would also imply an increase in futures market prices. As a consequence, the authors expect a

[14]We can talk about a backwardation situation when futures prices are lower than spot prices.

[15]We can talk about a contango situation when futures prices are higher than spot prices.

negative relationship between the two variables.
Among the proposed models there is also the introduction of the variables of spot returns and skewness, these are used to consider also different moments of the spot price, but their contribution to the model is proven to be almost irrelevant. The last analyses proposed are related to the use of some dummy variables, to study the CY on the basis of the Phase maturity of futures contract or directly on their year of maturity.
The main contribution achieved evidence that this spread is deeply related to the short-term rate, the market volatility and the amount of banked contracts.

1.4.2. Inconvenience Yield

Another important contribution in this field is given by Palao and Pardo (2021), whose article presents an analysis focused on the value of this negative spread, defined again as convenience yield. In particular, they consider the spread between ICE daily futures prices, which represent the EUA spot market, and ICE December EUA futures.
Again, the contango situation of the market is highlighted, since the combination of negative spread values and contango market allows investors to exploit arbitrage opportunities. The aim of the paper is to derive the main drivers of this negative spread.
To get a continuous time series of future prices and of spread values, the authors propose a building procedure using a rolling process with the last day criterion. In this, the switch in December contract price happens when the nearest December futures contract expires. From this series of prices one can derive the corresponding CY, obtaining once again a continuous time series.
The analysis proposes the use of a quantile regression, to analyse the different dependencies based on the quantiles of the dependent variable. The regressors selected are the returns in spot EUAs price, the intraday volatility in futures prices, two dummy variables to estimate the hedging and the speculative pressure in the market, the returns in MSCI European index, the Eurostoxx50 volatility index, the returns in IBOXX Euro Corporates AAA index and the Germany 10 year zero coupon yield curve.
The results obtained from the analysis evidence that the carbon related variables, like the spot price and its volatility are the main drivers of the spread, especially in case of low volatility and when the market is bearish. However, the model works better, and its explanatory power increases when financial variables are also considered.

1.4.3. Commodities Contribution

The last article taken as main reference during the analysis is the Alberola et al. (2008), which is one of the first studies on the relationship between carbon contracts spot price

and commodity related variables. This work focuses on the EUAs spot market in Phase 1 (2005-2007), which saw two structural breaks in 2006. These breaks were due to some changes in Europe's environmental policy, and after the report on verified emissions of 2005.
The aim of the authors is to find the main drivers of the carbon spot price, and to analyse its change when compared to commodity prices and temperature changes. The main variables used in the regressions are oil price, natural gas price, coal price and electricity price. For what concerns temperature variables, they are constructed as temperature averages of different regions in Europe.
The regression model is applied in three different time windows, in first place the whole Phase 1, then on the two subsets before and after the structural break of April 2006.
The results show evidence that the full Phase 1 analysis has different outcomes with respect to the analyses proposed on the sub-periods. In fact, it seems that in different time windows, different commodities get relevant, concluding that the main drivers in the spot price depend on the time period considered, and confirming that commodities can influence the spot market.

2 | Description of the dataset

This chapter describes the dataset used for the analysis, specifying the sources of the data and the relative characteristics.
All the datasets relative to prices and rates were downloaded from Refinitiv Eikon, one of the most widespread data providers of financial data, while the information regarding allowances allocations and verified emissions are retrieved thanks to the European Union Transaction Log, which records the information linked to compliance entities on verified emissions and allowances allocations. The only data obtained from a different source is the 3-month German rate, which is downloaded from Investing.com[1], one of the top three global financial websites.

2.1. Futures Database

As previously said, there are two main European Union Allowances exchanges, EEX and ICE, our study is focused on the latter one, being it the most liquid and actively traded futures market. For the analysis we obtained the spot prices and futures prices of the EUAs contracts as of January 13, 2009, or the first available date, for both exchanges. Then, after different checks on the daily traded volumes and open interests, values present in the dataset of futures contracts, and confirmed also in the recent literature (see e.g. Palao and Pardo (2021); Ellerman et al. (2016); ESMA (2022)), we decided to keep ICE as the reference market. To confirm this, as previously seen in Figure 1.3, one can see that the average volumes traded in ICE exchange are by far higher than the ones in EEX. Since ICE does not have an official spot price, the daily future on EUAs has been considered as proxy, similarly to Palao and Pardo (2021). These daily contracts have the same characteristics as the previously described futures on EUAs, but they have as delivery period the business day following their trade date.

[1]The reference website from which the data is obtained is: *https://www.investing.com/rates-bonds/germany-3-month-bond-yield-streaming-chart* .

The full list of ICE futures contracts used in the analysis is available in Table 2.1. In this we include the daily futures, used as spot prices, and the December futures, used to obtain the spread values.

List of Futures Considered

Ticker	Description	Currency	Delivery	Start time
0CFI2ZZ9^0	EUA December Future maturing on 14/12/2009, on ICE exchange	EUR	Physical	13/01/2009
CFI2Z0^1	EUA December Future maturing on 20/12/2010, on ICE exchange	EUR	Physical	13/01/2009
CFI2Z1^1	EUA December Future maturing on 19/12/2011, on ICE exchange	EUR	Physical	13/01/2009
CFI2Z2^1	EUA December Future maturing on 17/12/2012, on ICE exchange	EUR	Physical	13/01/2009
CFI2Z3^1	EUA December Future maturing on 16/12/2013, on ICE exchange	EUR	Physical	13/01/2009
CFI2Z4^1	EUA December Future maturing on 15/12/2014, on ICE exchange	EUR	Physical	13/01/2009
CFI2Z5^1	EUA December Future maturing on 14/12/2015, on ICE exchange	EUR	Physical	06/08/2010
CFI2Z6^1	EUA December Future maturing on 19/12/2016, on ICE exchange	EUR	Physical	06/08/2010
CFI2Z7^1	EUA December Future maturing on 18/12/2017, on ICE exchange	EUR	Physical	06/08/2010
CFI2Z8^1	EUA December Future maturing on 17/12/2018, on ICE exchange	EUR	Physical	06/08/2010
CFI2Z9^1	EUA December Future maturing on 16/12/2019, on ICE exchange	EUR	Physical	06/08/2010
CFI2Z0^2	EUA December Future maturing on 14/12/2020, on ICE exchange	EUR	Physical	06/08/2010
CFI2Z1^2	EUA December Future maturing on 20/12/2021, on ICE exchange	EUR	Physical	31/01/2017
CFI2Z2^2	EUA December Future maturing on 19/12/2022, on ICE exchange	EUR	Physical	31/01/2017
CFI2Z3	EUA December Future maturing on 18/12/2023, on ICE exchange	EUR	Physical	31/01/2017
CFI2Z4	EUA December Future maturing on 16/12/2024, on ICE exchange	EUR	Physical	31/01/2017
CFI2Z5	EUA December Future maturing on 15/12/2025, on ICE exchange	EUR	Physical	31/01/2017
CFI2Z6	EUA December Future maturing on 14/12/2026, on ICE exchange	EUR	Physical	17/12/2019
CFI2Z7	EUA December Future maturing on 20/12/2027, on ICE exchange	EUR	Physical	15/12/2020
CFI2Dc1^1	EUA Daily Future in Phase 2 (ends on 31/12/2012), on ICE exchange	EUR	Physical	13/01/2009
0CFI3Dc1^2	EUA Daily Future in Phase 3 (ends on 31/12/2020), on ICE exchange	EUR	Physical	02/01/2013

Table 2.1: For every contract the time series ends in its maturity date or on October 22, 2022 when the maturity is over said date. The frequency of the data has daily granularity and European holiday dates.

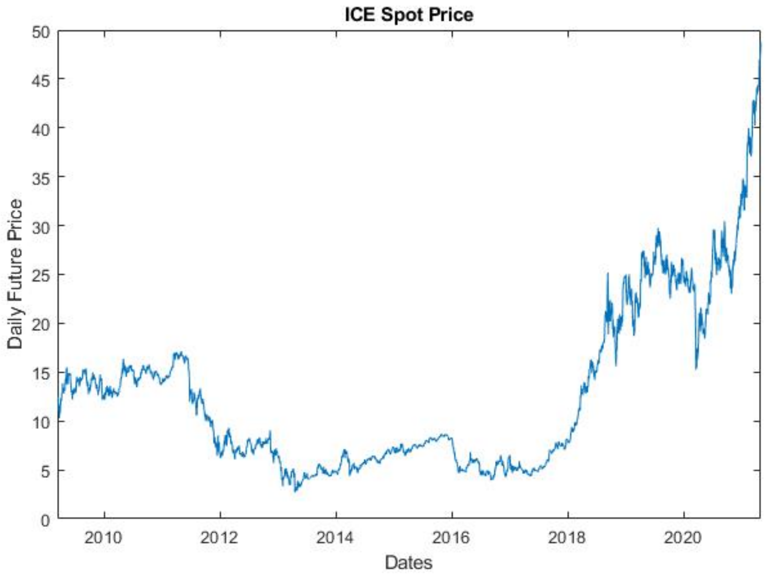


Figure 2.1: Graph of Daily Future Prices (spot price), with values expressed in euros. This is obtained stacking the vectors of daily futures of the two phases. Data Source: Refinitiv Eikon.

In the analysis, we restricted the study to futures with maturity in December as they are the most liquid contracts (Trück and Weron (2016); ESMA (2021); Swinkels and Yang (2022)). Confirmations in this sense come from Figure 2.2 and Figure 2.3 where it can be seen that the open interests of contracts with maturity outside December are negligible with respect to the ones maturing at the end of the year.

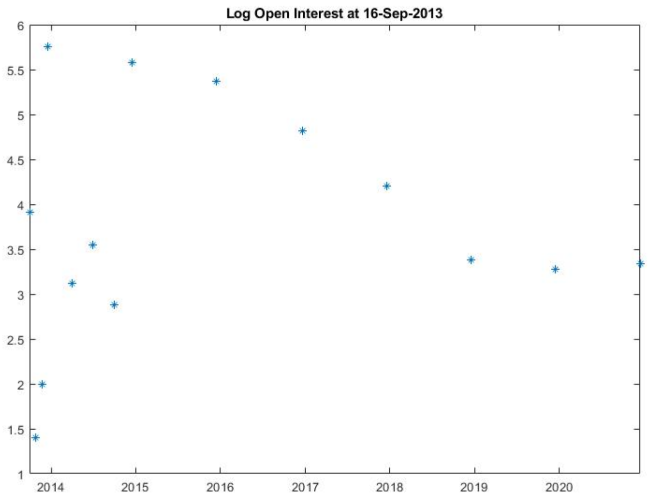


Figure 2.2: Graph of the logarithmic open interest for available contracts as of September 16, 2013, taken as a random sample. The y-axis represents the logarithm (base 10) of the open interest of traded contracts, while the x-axis represents the maturities of the contracts. Data Source: Refinitiv Eikon.

Some estimates report that more than 90% of the contracts traded during the year are represented by the futures maturing in December (Quemin and Pahle (2023); Ellerman et al. (2016)).

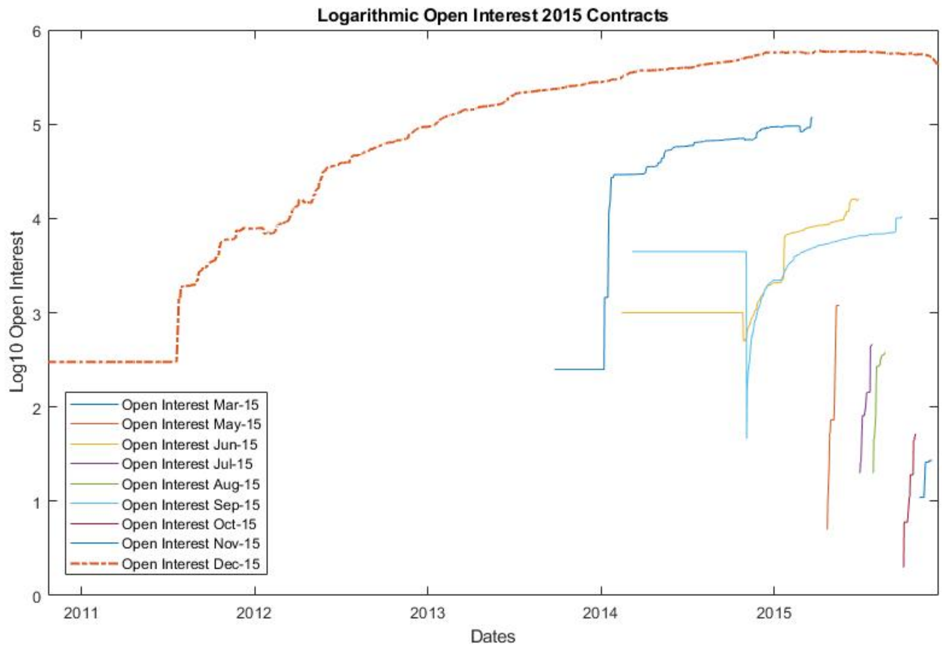


Figure 2.3: Graph of the open interest for 2015 futures contracts, with the y-axis representing the logarithmic open interest of the contracts and the x-axis representing the trading dates. The figure proves that the most important contract is the one maturing in December because it has the longest time series and the highest open interest. In particular, with respect to the December contract, September and June contracts have open interest almost two orders of magnitude lower, and March contract has at least one order of magnitude less, in best case scenario. The remaining contracts have more than three orders of magnitude as difference. Data Source: Refinitiv Eikon.

The reason behind the different order of magnitude in contract volumes, and open interest, stands behind the regulation rules imposed by the EU-ETS. Compliance entities have to monitor the emissions of CO_2 during the year and report them within December 31 every year, thus the last future allowing for physical delivery of EUAs before the submission of emissions at European level is the December contract, making it the most preferred one (Swinkels and Yang (2022)).
To complete on regulation rules, emissions have to be verified at national level, within March 31 of the following year so that the compliance entities can submit and surrender the needed amount of allowances not later than April 30[2]. Due to this fact, the March

[2]Further information can be found in *Monitoring, reporting and verification of EU ETS emissions: https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets/monitoring-reporting-and-verification-eu-ets-emissions_ en* (accessed on 10 March 2023).

future represents the second most liquid futures contract in the market, but the volumes traded are by far lower with respect to the December ones. Other mostly traded futures are the ones maturing in June and September, but they are even less liquid than the March ones. The contracts expiring in the remaining months are irrelevant as the volumes traded are insignificant when compared to the December futures' volumes, and data related to those kinds of contracts have history of three months maximum.

2.2. Allowances and Emissions Database

Another relevant variable considered in the study is the banked allowances estimate. This data is an indirect information that is estimated similarly to Trück and Weron (2016). We obtained the records relative to freely allocated allowances, verified emissions, values of offset and total allocated allowances, and from them we estimate the yearly banked allowances as the equation (2.1):

$$Banking(t) = Banking(t\text{-}1) + (FreeAlloc(t) - VerEmiss(t) + Offset(t)) \tag{2.1}$$

In the equation, t has yearly granularity, so we have data relative to the variables with yearly recurrence, moreover every value has unit of measure "equivalent tons of CO_2". Hence, $Banking(t)$ refers to the tons of CO_2 equivalent banked, $FreeAlloc(t)$ refers to the EUAs given for free that year, $VerEmiss(t)$ represents the verified emissions relative to that year, and $Offset(t)$ refers to the tons of CO_2 equivalents compensated with contracts that were not EUAs, being CERs and ERUs.
A relevant observation to add is that the possibility to bank contracts started in 2008 so previous values of this data are set to zero. Moreover, after 2012 no CERs and ERUs were used to compensate emissions so the value of offset goes to zero after that year.
This way to retrieve the data might work for the years before 2013, since from there the free allocations drastically dropped, so we propose another way to intend the *Banking* variable as equation (2.2):

$$Banking(t) = Banking(t - 1) + (TotalAlloc(t) - VerifiedEmiss(t)) \tag{2.2}$$

In which the $TotalAlloc(t)$ represents allocations derived both from free allocations and traded allowances, and it already considers the values of offset.
The data used for the extraction of *Banking* are obtained via the European Union Transaction Log (EUTL) which is the most reliable data source for information relative to allowances and emissions (Abrell (2021)). Figure 2.4 shows the data used in previous

equations, to give the reader a view of the phenomenon observed. While Figure 2.5 shows the yearly observations and the interpolated daily observations of the variable *Banking* when considering the total allocations, which is the preferred case. In Appendix A one can see the figure with free allocations, proposed by Trück and Weron (2016).

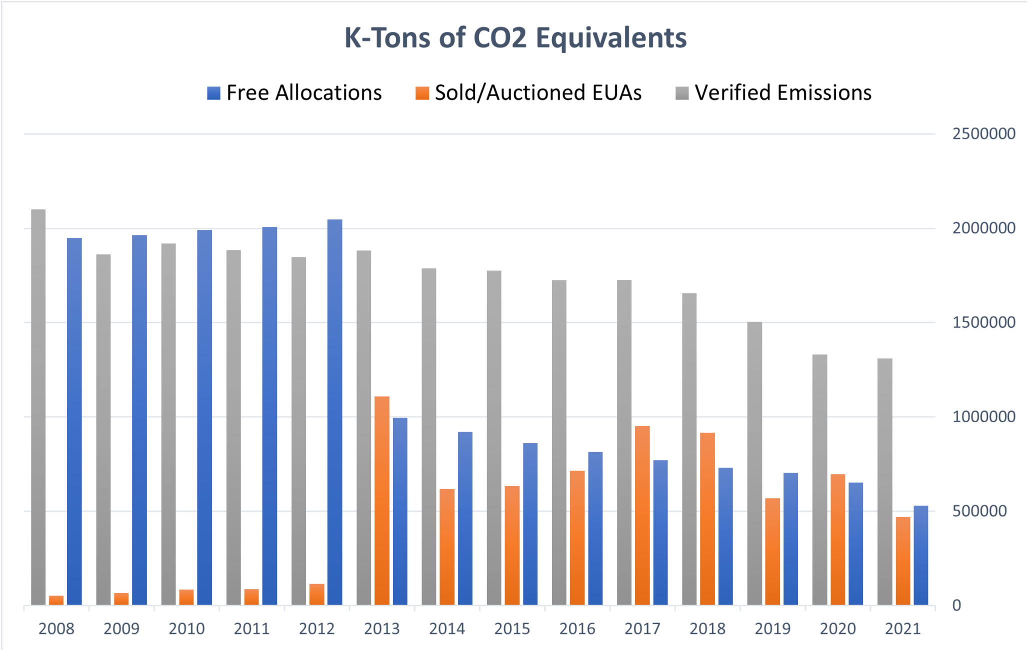


Figure 2.4: Graph of Free allowances allocated, Sold or Auctioned allowances and Verified emissions per year, with data measured in kilotons of CO_2 equivalent. Data Source: European Environment Agency.

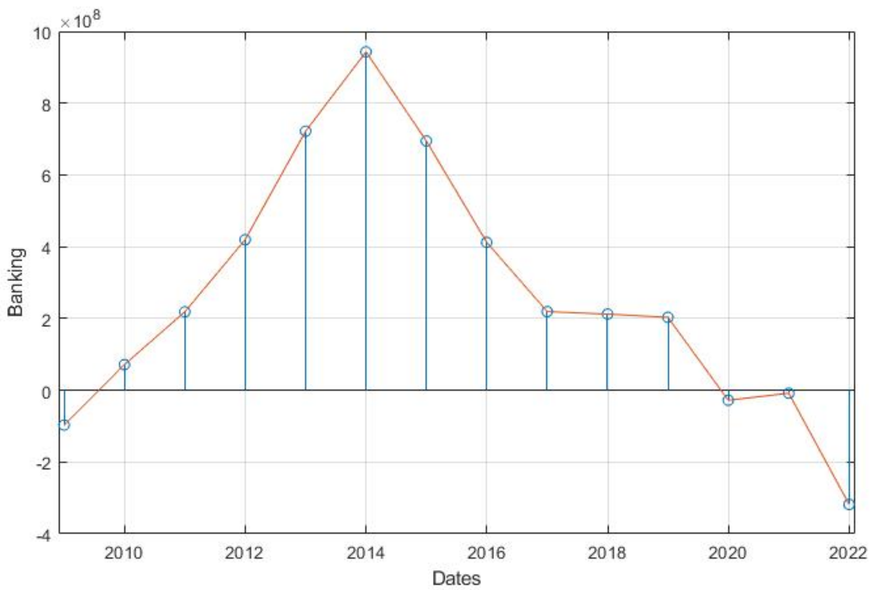


Figure 2.5: Graph of banked contracts derived using equation (2.2), with the unit of measure expressed in tons of CO_2 equivalent. Source: European Environment Agency.

3 | Methodology

In this chapter we describe the processes used in the analysis, presenting the formulas and regression models whose results are discussed later. After each model's presentation there is a brief explanation of the reason behind the variables' introduction.

3.1. Equations and terminology

As a starting point, the first variable to look for is the risk-free rate. For every date-time considered, these rates are computed as in the work of Baviera and Cassaro (2015) using the Overnight Index Swap (OIS) rate. When the time interval between t_0 and t_e is less or equal than one year, following the discounting curve bootstrap approach one can obtain the discount factors $B(t_0, t_e)$ as:

$$B(t_0, t_e) = \frac{1}{1 + \delta(t_0, t_e) R^{OIS}(t_0, t_e)} \tag{3.1}$$

In the equation, $R^{OIS}(t_0, t_i)$ represents the OIS rate settled at time t_0 lasting till time t_e, and $\delta(t_0, t_e)$[1] defines the time interval between t_0 and t_e. In case of maturity longer than one year one can bootstrap the needed discount factor, with the formula that slightly changes to:

$$B(t_0, t_i) = \frac{1 - R^{OIS}(t_0, t_i) \sum_{k=1}^{i-1} \delta_k B(t_0, t_k)}{1 + \delta_i R^{OIS}(t_0, t_i)}. \tag{3.2}$$

In which $R^{OIS}(t_0, t_i)$ represents the OIS rate settled at time t_0 and lasting till time t_i, δ_i defines the time interval between t_i and t_{i-1}, and $B(t_0, t_k)$ is the discount factor, previously calculated, between t_0 and t_k. As previously said, equation (3.1) is used when the time interval between t_0 and t_e is less or equal than 1 year, while the discount factor at longer maturities is computed via equation (3.2). After having retrieved the discount factors, it is easy to obtain the risk-free rates as:

$$r_{t_0, t_e} = r(t_0, t_e) = -\frac{log(B(t_0, t_e))}{\delta(t_0, t_e)}. \tag{3.3}$$

[1]To compute the time interval the time convention used is act/365.

To obtain the needed rates, we downloaded all the data available from March 13, 2009[2], up to end of October 2022, considering available contracts with maturities between 2 weeks and 30 years, including also 15, 18 and 21 months maturities. Once the rate is set, the spread between futures prices and spot prices can be easily computed as equation (3.4):

$$s_{t_0,T} = r_{t_0,T} - \frac{\ln(F_{t_0,T}) - \ln(S_{t_0})}{T - t_0}. \tag{3.4}$$

Again, following Trück and Weron (2016), the spread term is considered in negative term as it represents a convenience yield in their work. In this formula, $s_{t_0,T}$ defines the spread in t_0 for the future expiring in T, $F_{t_0,T}$ and S_{t_0} represent respectively the futures price and spot price of the good, and finally $r_{t_0,T}$ is the risk-free rate. As the rates obtained have monthly or yearly maturity, whenever a risk-free rate for maturity T is not available, one can linearly interpolate the two closest rates to get an estimation of the needed value. Again, the difference in time at the denominator represents the year fraction of the difference, calculated using the usual convention Act/365.

To give the reader a view of the phenomenon observed, we can see the market in contango situation, in Figure 3.1, and the spread observed for these contracts in Figure 3.2.

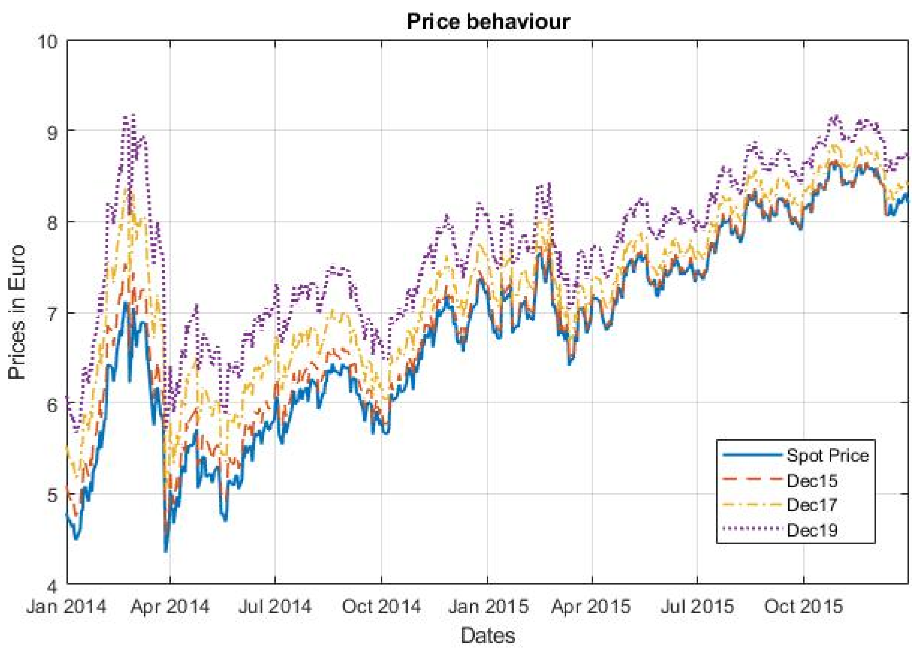


Figure 3.1: Graph of spot and futures contract prices, with the ICE daily future price shown in blue, the price of the futures maturing in December 2015 shown in orange, the price of the futures maturing in December 2017 shown in yellow, and the price of the futures maturing in December 2019 shown in purple. All prices are reported in euros. Data Source: Refinitiv Eikon.

[2]We choose this date as it is the first one in which data relative to all variables used are available.

Figure 3.2: Graph of the observed spread for contracts maturing at different times, with the spread curve for the future contract maturing in December 2015 shown in blue, the spread curve for the future contract maturing in December 2017 shown in orange, and the spread curve for the future contract maturing in December 2019 shown in yellow. All spreads are expressed in percentage.

As the figure shows, the spread becomes more volatile as contract maturity gets closer, while remains stable when its maturity is far from the considered date. In particular, if the term structure of the spread is considered, as in Figure 3.4, one can notice that the phenomenon is represented by a decreasing curve whose slope tends to go to zero as longer maturities are considered.

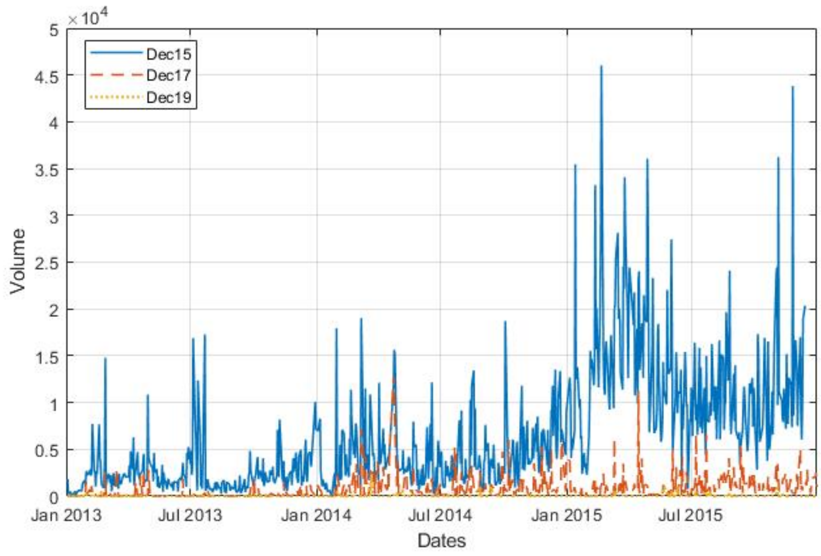


Figure 3.3: Graph of the observed volumes traded for contracts maturing at different times, with the volume curve for the future contract maturing in December 2015 shown in blue, the volume curve for the future contract maturing in December 2017 shown in orange, and the volume curve for the future contract maturing in December 2019 shown in yellow. All volumes expressed in lots*10^4. Data Source: Refinitiv Eikon.

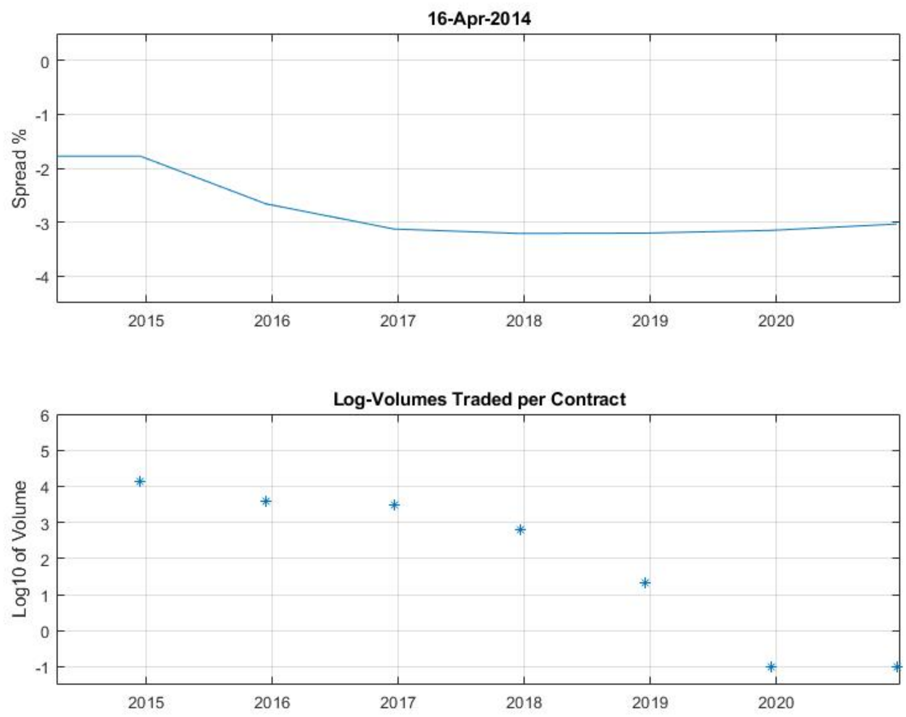


Figure 3.4: The top figure represents the term structure of the spread, expressed in percentage, as of April 16, 2014. The points considered are the observed spreads for different futures contracts. The bottom figure shows the logarithm of volumes traded on the day of consideration. If no trades occurred for the considered contract, a value of -1 is applied. Data Source of the volumes: Refinitiv Eikon.

3.2. Previous Models

In this section we present the models considered when conducting the study, taking as a basis the ones presented in Trück and Weron (2016). For the analysis a pooled ordinary least squares (OLS) estimation has been used, this means that all the observations of each contract are considered in the same regression. Hence, a vector of spreads S_t is created using the values of spreads $s_{t_0,T}$ of all the available contracts.
The first model considered, from now on Model I, is presented in equation (3.5).
The variables selected for the regression are the 3-months German rate ($RATE_t$), banking amount of contracts ($BANKS_t$) and the volatility of spot price returns (VOL_t). The first one is selected as a proxy for short-term rates, as the spread is related to risk free rates one can expect a high level of dependency between the two variables. As a short-term rate we select the German rate, while Trück and Weron (2016) choose the ECB rates previously described. The reason behind this change is to have a rate that is actually observed on the market and not a rate derived by the info-provider.
Banking, intended as contracts stored, is selected as generally for commodities the spread represents the benefit of holding the good now. Since we have yearly data relative to banking, Trück and Weron (2016) suggest a linear interpolation to obtain the value of banked contracts in a specific date. For what concerns the spot returns volatility a GARCH(1,1) model has been applied to the returns of spot prices, to get the spot volatility[3]. This value is taken into consideration since in periods of high volatility hedging demand increases, consequently futures prices and spread values should be influenced.

$$S_t = \beta_0 + \beta_1 RATE_t + \beta_2 BANKS_t + \beta_3 VOL_t + \epsilon_{i,t}. \tag{3.5}$$

The following model, Model II, introduces two additional variables. These are the most recent spot return (r_t) and the skewness in the spot market ($SKEW_t$). The latter is obtained via equation (3.6) as a process similar to a 20-day moving average.

$$SKEW_t = \frac{1}{k-1} \sum_{i=1}^{k} \frac{(r_i - \bar{r})^3}{\sigma_t}. \tag{3.6}$$

In the equation, k is selected equal to 20, r_i represents the the spot return at time i, $\bar{r}$ denotes the average of spot returns in the last k days, and σ_t defines the standard deviation of the returns at time t.

[3]Trück and Weron (2016) use both a GARCH(1,1) and an exponential moving average model to get the volatility, we decide to use the GARCH(1,1) model, that seems to catch better the high peaks of volatility.

The two variables are introduced as they might have an impact on risk premiums in the futures markets (Kumar and Trück (2014)). So, the second model extends the first one adding these market variables and is defined as equation (3.7).

$$S_t = \beta_0 + \beta_1 RATE_t + \beta_2 BANKS_t + \beta_3 VOL_t + \beta_4 r_t + \beta_5 SKEW_t + \epsilon_{i,t}. \tag{3.7}$$

Since the spread tends to be lower for futures with distant maturity, a dummy of phase is introduced (d_{phase}). In fact, contracts that close in the same phase tend to have less pronounced spread, when compared to the ones maturing in the following phase. A dummy variable is then introduced to take into consideration this fact. Model III is then defined as equation (3.8).

$$S_t = \beta_0 + \beta_1 RATE_t + \beta_2 BANKS_t + \beta_3 VOL_t + \gamma d_{phase} + \epsilon_{i,t}. \tag{3.8}$$

To further investigate the relations among variables a new model, Model IV, is implemented with the introduction of the phase dummy to Model II.

After having divided the spreads splitting them on the basis of their phase's maturities, to analyse the differences on contract maturity basis a final model, Model V, is introduced. In this, a dummy variable for each spread's maturity is proposed (d_i), the full formulation is given in equation (3.9).

$$S_t = \beta_0 + \beta_1 RATE_t + \beta_2 BANKS_t + \beta_3 VOL_t + \sum_{i=2}^{n} \gamma_i d_i + \epsilon_{i,t}. \tag{3.9}$$

In the model, i starts from 2 to avoid the dummy variable trap and n represents the number of futures contracts considered. As done before, a modification of the model, Model VI, is proposed adding the variables relative to returns and skewness to Model V.

3.3. Model Improvements

In this section we introduce the models used to further investigate the behaviour of the spread. The first models are similar to the previous six, while the latter ones have major differences, since they see the introduction of the new variables proposed. However, Model I is always considered as a starting point for the models proposed.

The first extension of the previous models sees the introduction of the variable of volume traded ($VOLUME_t$). This variable is expected to be significant because when the volume of trade increases, usually in the last year and a half before maturity, the spread tends to go towards the zero value, as can be observed in Figure 3.2 and Figure 3.3. In equation (3.10) it is shown the modification of Model I with the volume variable.

$$S_t = \beta_0 + \beta_1 RATE_t + \beta_2 BANKS_t + \beta_3 VOL_t + \beta_4 VOLUME_t + \epsilon_{i,t}. \tag{3.10}$$

As all the other models would be similar but with the introduction of the new variable it would be redundant to present them all once again.

A second study that we perform is the analysis of spread values considering observations in which the volume traded exceeds a certain threshold. This "filtering" is done to select only the observations whose price is determined as a weighted average of trading prices. As discussed in the introduction, ICE settlement prices are not the weighted average of traded prices when the volume traded in the last ten minutes of the day is less than 100 lots. Since the trading session is composed of ten hours, we decide to select as a filter the condition: $\frac{Volume}{60} \geq 100$, supposing that the trades are equally spread during the trading session. We also implement some robustness checks with other thresholds, in particular we present the results with a tighter threshold $\frac{Volume}{80} \geq 100$ and a looser one $\frac{Volume}{40} \geq 100$.

To further investigate the behaviour of the spread in EUAs market we shift the focus on the commodity world. As Alberola et al. (2008) found a relationship between carbon spot price and commodities related variables, we try to extend their study analysing the relation with spread values. To do so, four commodities are selected: Brent price for oil, TTF monthly future price for gas, the German baseload power price for energy and API2 CIF for coal futures price. These data have been downloaded from Refinitiv Eikon, and before their usage all prices were converted in euros. Table 3.1 provides the data relative to the commodities considered.

List of Commodities Considered

Ticker	Description	Currency	Delivery	Start time
$BRT-$	Brent crude spot price, free on board	USD	Physical	13/01/2009
$TFMBMc1$	TTF Monthly future, in ICE exchange	EUR	Physical	12/03/2010
$TRDEBD1$	TRPC electricity Germany daily baseload	EUR	Physical	13/01/2009
$TRAPI2Mc1$	API2 CIF future price, in ICE exchange	USD	Physical	13/01/2009

Table 3.1: For every contract the time series ends on October 22, 2022. The frequency of the data has daily granularity. Every price has been converted in euros before being downloaded.

To create the model, we focus on the relationship between the spread and the price of the commodities (P_i) and its volatility (σ_i^2), as in Alberola et al. (2008). Hence, the reference model is presented in equation (3.11).

$$S_t = \beta_0 + \beta_1 RATE_t + \beta_2 BANKS_t + \beta_3 VOL_t + \beta_4 VOLUME_t + \sum_{i=1}^{n} \gamma_i P_i + \sum_{i=1}^{n} \rho_i \sigma_i^2 + \epsilon_{i,t}. \tag{3.11}$$

In which i is the index used to consider all the n commodities, P_i represents the commodity's price and σ_i^2 its price's volatility, obtained once again applying a GARCH(1,1) model, while γ_i and ρ_i represent the coefficients of the regression for the considered commodity.

Also in this case a variation of the model is proposed, with the application of the volume filter, to study the relationship between spread and commodity variables when the price is artificially created and when the settlement price is the true weighted average.

A further analysis is presented on the basis of the free allowances reduction in 2013. In fact, from the beginning of Phase 3, more companies have to obtain allowances via secondary market or via auctions, with power generators suffering the most for this issue as their free allocations dropped to zero. So, the analysis focuses on the credit spread between CO_2 emitters bonds and the risk-free rate, defined Z-spread (Z_t), compared to the EUAs spread. To do so we propose the model in equation (3.12). The analysis on this model is presented with and without the use of the volume filter.

$$S_t = \beta_0 + \beta_1 RATE_t + \beta_2 BANKS_t + \beta_3 VOL_t + \beta_4 VOLUME_t + \mu_1 Z_t + \epsilon_{i,t}. \tag{3.12}$$

In the model, Z_t represents the average of the 2 years Z-spreads derived from companies with highest GHG emissions. Z-spreads are obtained from benchmark issues, i.e. contracts with issue amount greater or equal than 500Mil. After having selected the companies with high gas emissions and after having filtered the coupon bonds, we obtain the Z-spread using the coupon bond prices (see e.g. Baviera et al. (2021)) via equation (3.13):

$$Z(t_0, T) := -\frac{1}{T - t_0} \ln \frac{\overline{B}(t_0, T)}{B(t_0, T)}. \tag{3.13}$$

Where $\overline{B}(t_0, T)$ represents the discount factor obtained from the quoted bonds and $B(t_0, T)$ the risk-free rate discount factor. The day-count convention when computing $T - t_0$ is the standard Act/365. In particular, the value of $\overline{B}(t_0, T)$ can be obtained from the bond prices by inversion of equation (3.14):

$$C_{t_0,T} + Acc = \sum_{i=1}^{n-1} c\delta_i \overline{B}(t_0, t_i) + (100 + c\delta_n)\overline{B}(t_0, T). \tag{3.14}$$

In the equation, $C_{t_0,T}$ is the coupon bond clean price, Acc represents the accrual term, c the coupon and δ_i the time interval between coupon payment days, where δ_1 is from the last observed payment date up to the next one.

Finally, after having seen that adding the lagged spread S_{t-1} to the regression the adjusted R^2 shoots over 75% in both filtered and non-filtered case, we try to create an auto-regressive model. To do so, inspired by Palao and Pardo (2021), we create a time series for the spread, obtained rolling the front December contract for the whole Phase 3. This time series is obtained considering the spread of only the nearest front December contract. For example, in between January and December 2013 the front December contract is the one expiring in December 2013, then at its maturity the front future becomes the one maturing in December 2014, this procedure is iterated till the time series for the whole Phase 3 is created. After having checked that the series is autoregressive with an auto-correlation function (ACF) test and a partial auto-correlation function (PACF) test, stationarity is verified by the use of an Augmented Dickey–Fuller (ADF) test. The final model applied is represented in equation (3.15):

$$S_t = \beta_0 + \beta_1 RATE_t + \beta_2 BANKS_t + \beta_3 VOL_t + \beta_4 VOLUME_t + \mu_1 Z_t + \mu_2 S_{t-1} + \epsilon_t. \tag{3.15}$$

Since the reduction of free contracts impacted only Phase 3 and since the Z-spread is bond to this event, we present the last two models only in Phase 3.

To sum up, in this chapter the main components and equations used in the study are introduced, moreover the models and variables used are described and explained. In the following chapters, the results obtained using these models are presented.

4 | Phase 2 Analysis

In this chapter we present the analyses based on the models described in Chapter 3. In particular, the chapter is divided in a fist part, which focuses on the replica of the Trück and Weron (2016) study, and a second part devoted to new models proposed. The whole study is conducted as in Trück and Weron (2016) on the time window of Phase 2 which goes from 2008 to the end of 2012. As specified in Chapter 2, the first date on which all the data are available is March 13, 2009, hence we analyse Phase 2 starting from this date.

4.1. Replica of the article

In this section the replica of the cited article is presented. Before diving into the analysis a clarification has to be made, the dataset used in the study is not the same as Trück and Weron (2016). In fact, we consider as spot price the daily futures prices since the spot price described in the article is not available, in addition the risk-free rates used to extract the spread, named Convenience Yield in the article, is different since they use ECB quotes relative to AAA-rated bonds, while the ones chosen in the current analysis are derived from OIS rates, moreover for the 3-month short-term rate, the article uses the 3-month rate derived from ECB quotes for AAA-rated bonds while we prefer to use the 3-month German government bond yield as it is a data directly observed on the market and is not subject to transformations, as the one used by Trück and Weron (2016). Another difference is on the starting point of the dataset, while Trück and Weron (2016) start from April 8, 2008, our starting point is in March 13, 2009. Finally, we consider all futures till maturity December 2017.

This first analysis is presented for Phase 2, whose time span is between 2008 and 2012. Our results show very similar evidences. Figure 4.1 gives a visible representation of the phenomenon analysed, presenting the spreads for futures maturing in Phase 2.

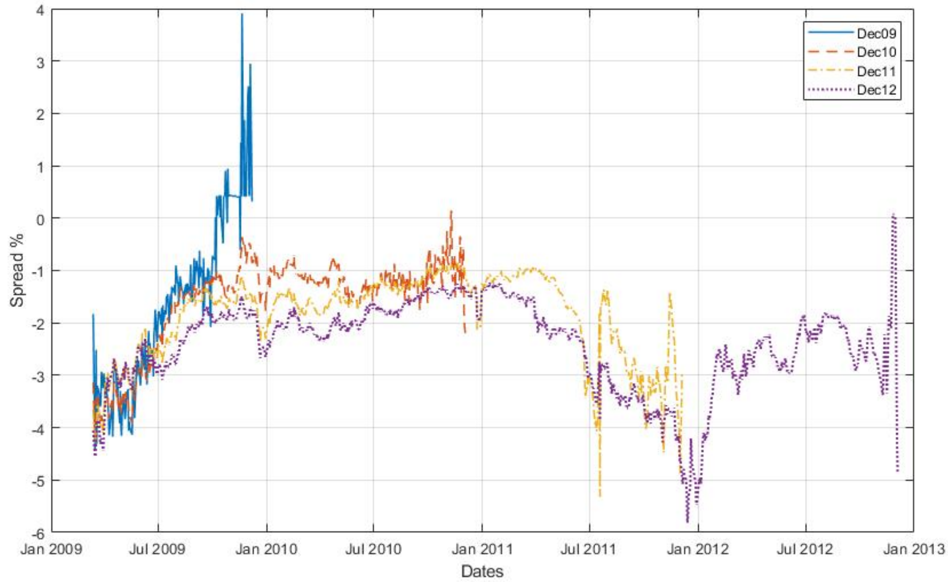


Figure 4.1: Graph of spread values expressed in percentage for futures contracts maturing in Phase 2.

We start with Model I (3.5), in which we consider the variables 3-month German rate, presented in Figure 4.2, as a proxy for the short-term rates, the interpolated banking as an estimate for the reserves of allowances (as in the work of Trück and Weron (2016)) and the volatility of spot returns, obtained applying a GARCH(1,1) model to the time series, to see the effect of market volatility.

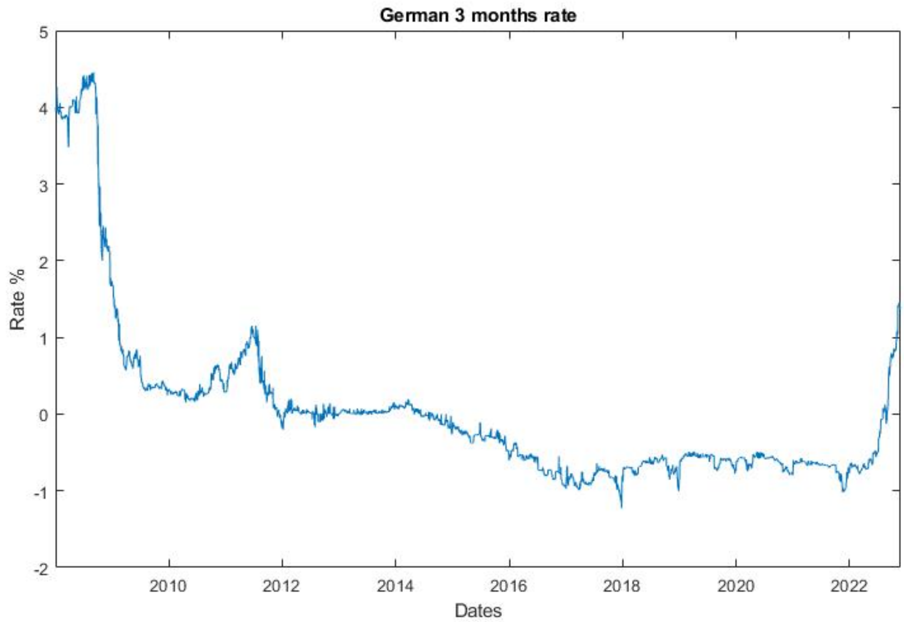


Figure 4.2: Graph of the 3-month German rate expressed in percentage. Data Source: Investing.com .

As shown in Table 4.1, the $adjusted - R^2$ slightly drops in all considered cases compared with the ones in Trück and Weron (2016), however, the level of significance of the variables and their signs are the same, pointing out that the change in variables influences the explanatory power of the models, but not the relevance of the variables. As Trück and Weron (2016) suggest, the 3-month rate is important as the spread depends directly from rates, from this the positive sign in the regression coefficient; banking are important data as they simulate the level of storage of the good and the negative sign in the regression is justified by the fact that the surplus of good lowers the benefit of holding it, hence the negative relation; while volatility is expected to have a negative sign as more uncertainty should increase future prices and consequently the spread.

Moving to Model II (3.7) for the first three variables the same considerations on Model I hold, while for what concerns the returns variable we reach the same result of the article, meaning we find a positive and significant relation with the spread, while the skewness variable presents a p-value lower than the one observed by Trück and Weron (2016). The reason behind that might stand in the change of dataset, in particular the change in the spot price. In fact, selecting the daily future price as the spot price automatically changes the returns and the returns' volatility, hence we can expect different results from the ones obtained by Trück and Weron (2016). However, a similar, and not significant, increase in $adjusted - R^2$ is visible, just like in the article, meaning that the introduction of the variables is not impactful in the model.

Considering Model III (3.8), the introduction of a variable related to the futures maturity phase has a remarkable impact in the regression, increasing the $adjusted - R^2$ by approximately 10% with respect to the same regression without the dummy of phase. The same result is obtained when the dummy is introduced in Model II, generating Model IV.
In both cases the relation with the variable is negative, suggesting that futures maturing in the successive phase present, on average, a higher level of spread when considering the variable in module.

After having verified that the phase dummy is relevant in the regression, an extension of the variable is made, introducing a dummy for each contract maturity. As observed for the previous case, the dummies are extremely important for contracts maturing after the end of Phase 2, emphasizing the fact that the spread, for contracts with distant maturities, is more pronounced.

Model	*I*	*II*	*III*	*IV*	*V*	*VI*
Intercept	-0.0174***	-0.017***	-0.0142***	-0.0138***	-0.0133***	-0.0122***
RATE	0.0024***	0.0031***	0.0042***	0.0052***	0.0042***	0.0052***
BANKS	-0.0308***	-0.0329***	-0.0204***	-0.022***	-0.0204***	-0.0216***
VOL	-8.1522***	-7.859***	-7.845***	-7.618***	-7.849***	-7.632***
r_t		0.0141***		0.0147***		0.147***
SKEW		0.3995***		0.5679***		0.573***
Phase3			-0.0111***	-0.0115***		
2010					-0.0029	-0.001
2011					-0.0007	-0.0017*
2012					-0.0014	-0.0024**
2013					-0.012***	-0.0128***
2014					-0.0131***	-0.0143***
2015					-0.0132***	-0.0145***
2016					-0.0116***	-0.0129***
2017					-0.0101***	-0.0115***
R-adj	50.6	52.6	62	64.4	62.3	64.8

Table 4.1: Table of results obtained from regressions using Models I-VI. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

4.2. New models

This section presents the new additions to the previously discussed models. These comprise the regressions involving volume and commodity variables. Moreover, a filter is implemented to study only prices deriving from ICE settlement procedure.

4.2.1. Volume analysis

As previously presented, after having replicated the models used in the article the study shifts its focus on looking for those variables that can help to increase the explanatory power of the model. As a first improvement, the volume variable is added to all previous models.

This variable increases the $adjusted - R^2$. Reason behind this fact is that when maturity comes closer the spread value tends to reduce towards zero, moreover as maturity approaches the number of lots traded increases, as can be seen in Figure 4.3, hence the two variables are in close contact and should present a positive relationship.

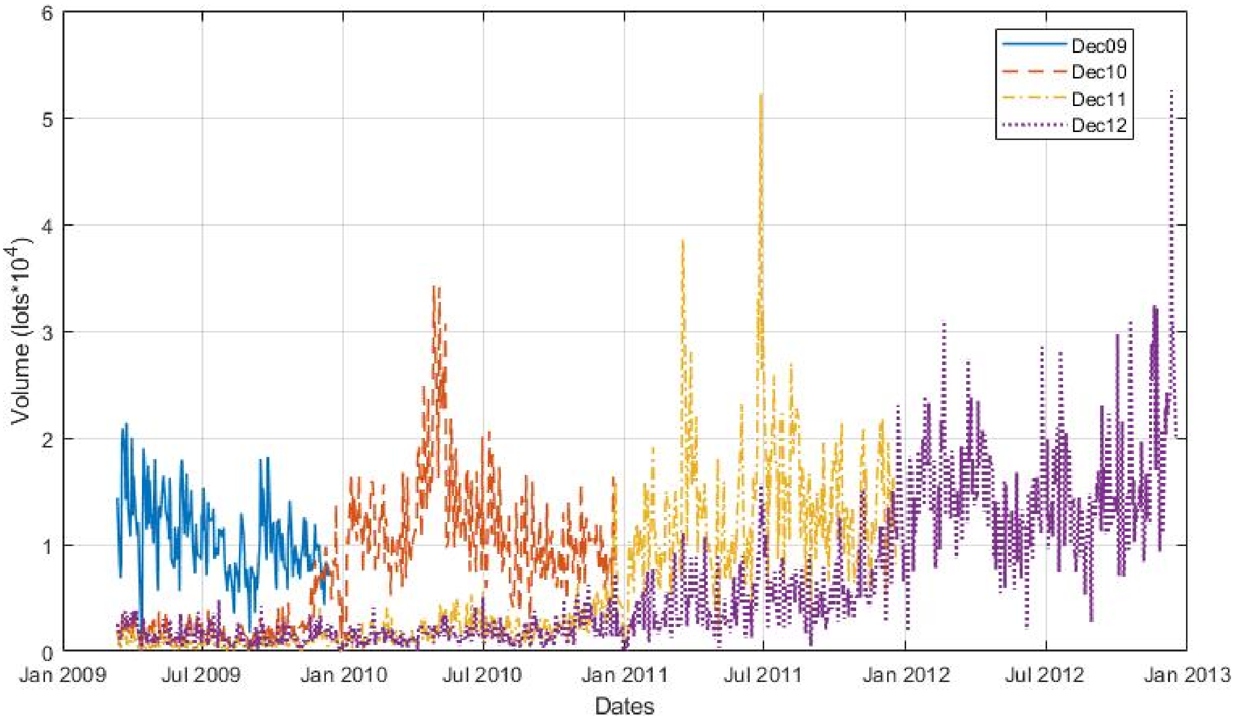


Figure 4.3: Graph of traded volumes for futures contracts maturing in December 2009, December 2010, December 2011, and December 2012, with the values expressed in 10,000 lots. Data Source: Refinitiv Eikon.

From Table 4.2 can be seen that the introduction of the volume variable is particularly significant for the first two models, while for the later ones the $adjusted - R^2$ is almost constant after the introduction of the variable. The reason behind this is that the phase dummy variable identifies all the contracts whose maturity is later than December 2012, those contracts have volume traded close to zero in Phase 2, since the increase in trading volume occurs almost a year before maturity, so the two variables catch similar information. Confirmation for this come from the fact that the coefficient for the volume variable decreases as we introduce the dummies of phase and maturity.

Model	*I*	*II*	*III*	*IV*	*V*	*VI*
Intercept	-0.191***	-0.0187***	-0.0154***	-0.0151***	-0.016***	-0.015***
RATE	0.0023***	0.0031***	0.0039***	0.0048***	0.0034***	0.004***
BANKS	-0.0326***	-0.0347***	-0.023***	-0.025***	-0.025***	-0.026***
VOL	-8.41***	-8.25***	-7.995***	-7.809***	-8.037***	-7.85***
VOLUME	0.79***	0.817***	0.2858***	0.3***	0.345***	0.34***
r_t		0.0127**		0.0141**		0.0141**
SKEW		0.655***		0.629***		0.616***
Phase3			-0.009***	-0.0093***		
2010					0.001	0.0001
2011					0.0018*	0.0006
2012					0.0013	0.0002
2013					-0.0076***	-0.0089***
2014					-0.0085***	-0.0098***
2015					-0.0079***	-0.0094***
2016					-0.0062***	-0.0077***
2017					-0.0047***	-0.0062***
R-adj	58.1	60.5	62.6	65.1	63	65.4

Table 4.2: Table of results obtained from regressions using Models I-VI with the addition of the volume variable. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

4.2.2. Filter Analysis

After having considered the volume variable, one can observe that the majority of the prices in the time series might not be market prices according to the regulation of ICE, since the minimum volume needed to have a weighted average price is 100, and for the observations before one year to maturity the volume traded is close to zero. So, the following step in the study is the proposal of a filter which limits the observations to the ones whose volume traded is superior to a certain threshold.
This threshold is set to be $\frac{Volume}{60} \geq 100$, due to the ICE settlement procedure proxy. Since this limit value is chosen by us, to get confirmation of the analysis we propose the same study on a more relaxed and a tighter threshold, namely, $\frac{Volume}{40} \geq 100$ and $\frac{Volume}{80} \geq 100$. From Figure 4.4 one can see the different thresholds used.
The results obtained with the middle threshold are reported below while the ones deriving from the other thresholds, used to check the robustness of the test, are reported in Appendix B.

After the implementation of the filter and the application of the model, the first thing to notice is that the $adjusted - R^2$ almost halves for each model.
Since the dummy of phase and the maturity dummies seem to capture the same phenomenon, we decide to present only the dummy of phase case. Moreover, after having seen that the returns and skewness variables are not significant at 5% level, and since the $adjusted - R^2$ does not increase much with their introduction, Models IV and VI are discarded from the current study.

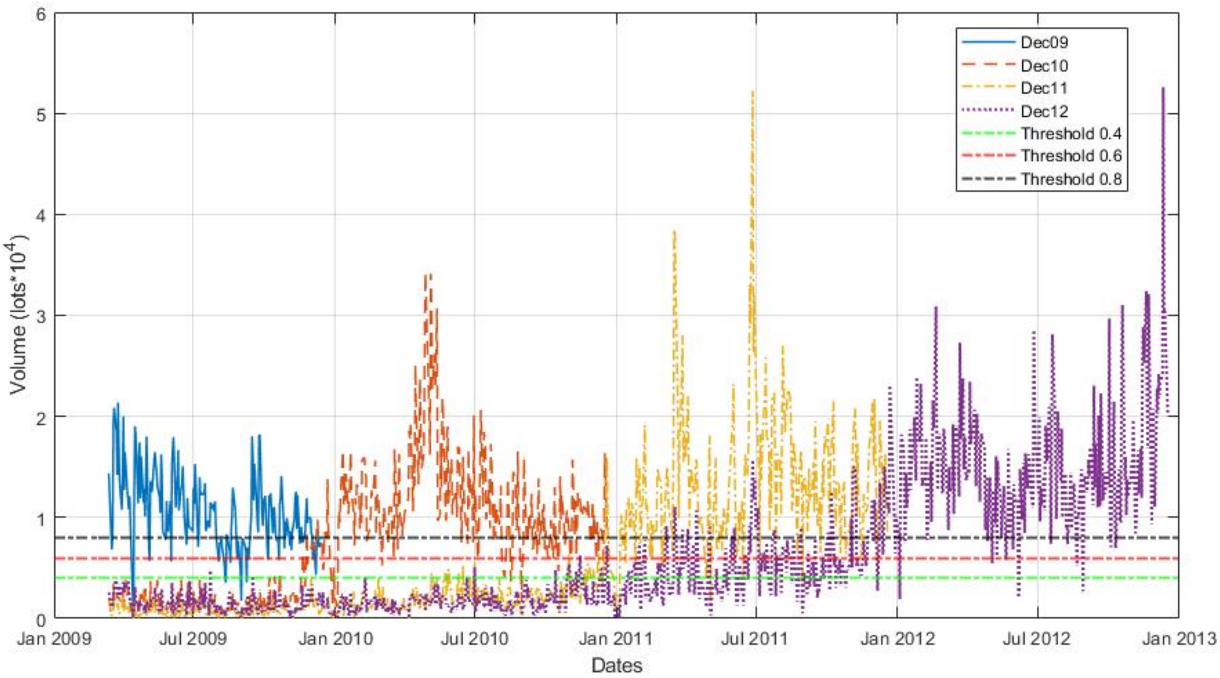


Figure 4.4: Graph of traded volumes for futures contracts maturing in December 2009 (in blue), December 2010 (in orange), December 2011 (in yellow), and December 2012 (in purple), expressed in 10,000 lots. The red dotted line represents the ICE proxy volume threshold, while the black and green lines represent the consistency check thresholds. Data Source: Refinitiv Eikon.

From these results the most relevant information to grasp is that the 3-month rate coefficient assumes negative sign and becomes not significant, at 1% level, in the regression, a result that is in contrast with Trück and Weron (2016) analysis.
Moreover, while the coefficients of banking and volatility variables behave the same way, the spot return, skewness, and volume variables lose their explanatory power and their relevance. Reason behind these changes might stand in the low necessity of contracts from compliance entities. In fact, as most contracts are obtained via free allocations, the market might lose the dependency from financial variables, such as the short-term rate.
Finally, looking at the Phase3 dummy, one can see that its introduction increases the

Model	I	I	II	II	III	III
Intercept	-0.008***	-0.009***	-0.007***	-0.008***	-0.008***	-0.008***
RATE	-0.003	-0.003	-0.003	-0.003	-0.0036*	-0.0035*
BANKS	-0.026***	-0.026***	-0.028***	-0.029***	-0.0202***	-0.0199***
VOL	-9.05***	-9.234***	-8.525***	-8.756***	-8.785***	-8.662***
VOLUME		0.099		0.113		-0.062
r_t			0.024	0.0232		
SKEW			-0.057	0.0031		
Phase3					-0.011***	-0.0115***
R-adj	28.8	28.8	30.1	30.1	31.5	31.4

Table 4.3: Table of results obtained from regressions using Models I-III on the filtered dataset. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

explanatory power of the model, but the rise is not as relevant as before. This is due to the limited number of observations relative to the next phase, as the majority of their volume is close to zero in Phase 2.

4.2.3. Commodities Analysis

The last model extension for Phase 2 proposes the analysis of the spread between futures and spot price of EUAs introducing variables relative to commodities. In Alberola et al. (2008), the authors focus on the analysis of carbon spot price with respect to commodities related variables. In the current research we implement a regression on spread against the evolution of commodities prices and volatilities to verify if a relation between the variables exists.

In this part of the analysis, the commodities considered are two: Oil and Power.As the variables related to coal and gas have time series not long enough to cover all the dates considered for allowances data.

The analysis is proposed on the dataset with and without the use of the volume filter. Before running the regressions, a multicollinearity analysis is made via the use of VIF, whenever the value of VIF exceeds 10 the variable is considered highly correlated with the others, so it is deleted from the regression. After the removal of these variables a p-value analysis is made to delete the variables which are not relevant at 1% level. Table 4.4 shows the results before and after the clearing of variables.

In Table 4.4, we observe that using the volume filter, results are rather different from the ones obtained without its use. When all data are considered, we can see an improvement in $adjusted - R^2$ of almost 2% with respect to the model without commodities seen in Table 4.2. Hence, at least for Phase 2 one can state that variables related to commodities

Filter	*Case No Volume Filter*		*Case Volume Filter*	
	Before Clearing	After Clearing	Before Clearing	After Clearing
Intercept	-0.027***	-0.018***	0.0004	0.0008
RATE	0.004***	0.0033***	0.005*	0.005*
BANKS	-0.0206***	-0.0269***	-0.0274***	-0.0275***
VOLS	-7.317***	-7.72***	-7.97***	-7.73***
VOLUME	0.891***	0.858***	0.196*	
Oil Price	0.0002***	-0.0001***	9.00E-05	
Oil Volatility	-3.00E-06***		-6.00E-07	
Power Price	5.00E-05**	6.00E-05***	-0.0002*	
Power Volatility	7.00E-07**	9.00E-07***	-3.00E-06***	-4.00E-06***
R-adj	61.1	59.9	32.5	32.2

Table 4.4: Table of results obtained from regressions with commodity models before and after clearing variables due to high multicollinearity or high p-value. On the left side are the regressions obtained considering all the observations, and on the right side are the results obtained considering observations with a volume traded over the threshold of 6000 lots. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

are not relevant for the spread analysis.
The same considerations hold when considering the filtered case, in which only the power price's volatility seems to survive the filtering procedure due to VIF and p-values. Moreover, a noticeable fact is that the coefficient of power volatility changes sign when the volume filter is applied, the same thing that happened in the RATE coefficient in previous cases. This confirms that settlement prices deriving from a weighted average of trading prices behave differently from non-filtered prices.

The main results obtained from the replica of the article and the extension of the presented model are that the variables of rate, banking and spot returns volatility are the main drivers of the spread, especially when considering all the available observations.
Moreover, volumes have importance in the analysis, in fact from them one can understand the liquidity of the futures. This data was previously caught with a dummy, but considering the raw data is far better.
Additionally, results change when considering the spread values for contracts whose volumes respect a certain threshold, so prices derived from non-liquid observations give different information with respect to the liquid ones.
Finally, the use of commodity related variables is useful especially when considering the whole dataset, but the improvement of the model, at least for Phase 2, is not significant.

5 | Phase 3 Analysis

In this chapter we consider Phase 3, i.e. the time window considered goes from 2013 to 2020. This phase saw the reduction to zero of free allowances allocated to power generators. Due to this fact, the analysis conducted on Phase 3 is considered more relevant than the previous one.
In this chapter, we apply the models introduced before and try to study the behaviour of the spread when compared to the Z-spread of GHG emitters, moreover the spread is analysed as an autoregressive process. Finally, a part devoted to the analysis of options is presented.

5.1. Extension of basis models

The first thing done is the extension of the models presented by Trück and Weron (2016) to Phase 3. This change of the time window is done to check if the results found before still hold, since the shift in phase took some market changes with itself. This dataset substitution is also used to verify the robustness of the results previously found.
After this brief introduction we can proceed to analyse the results obtained from the regressions. In this case the data considered are the ones relative to futures contract maturing after January 2013, so the analysis considers futures with maturity between December 2013 and December 2027, even if for this last contract the number of observations is so low that can be considered negligible and can be excluded from the analysis.
As previously done, we focus on one futures contract per year, the one maturing in December, as it represents the most actively traded futures.

Figure 5.1 presents the spread between the EUA daily futures and December futures prices for contracts at different maturities. As for Phase 2 spreads, also in this case we can observe that the spreads tend to converge to zero as maturity approaches and stay lower when the contract maturity is distant.

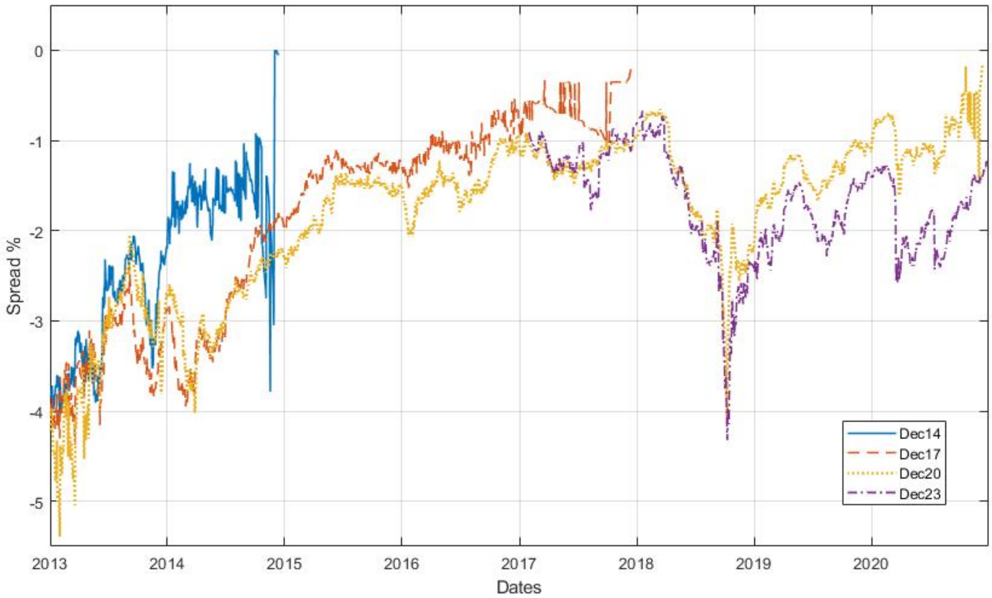


Figure 5.1: Graph of the spread obtained considering different futures contracts maturing in: December 2014 (in blue), December 2017 (in orange), December 2020 (in yellow), and December 2023 (in purple). The spread is expressed as a percentage.

The presented results are obtained considering data starting from January 2013, but to have a robustness check, in Appendix B one can see the results obtained considering the whole time series for contracts maturing in early stages of Phase 3, this means that for futures with maturity Dec 2013, Dec 2014 and so on, we consider the observations with date even before January 2013.

In Table 5.1 we present the results derived from the regression models introduced before, starting from the ones in the replicated article. The table does not have the introduction of volume variable and the observations are not filtered with the volume criterion, as this part is devoted just to the extension of the first study to a different time window.

We can observe results that are similar to the ones obtained in Chapter 4. In fact, all the significance levels are identical for the first four models, while when considering the maturity dummies, all variables seem to have relevance, suggesting that there is not homogeneity in the spread between the contracts. It is also important to notice that, when the contract maturity is higher than December 2020 the coefficient of the dummy stays almost equal for every contract, the same characteristic observed in Phase 2.

Model	*I*	*II*	*III*	*IV*	*V*	*VI*
Intercept	-0.027***	-0.027***	-0.019***	-0.019***	-0.007***	-0.007***
RATE	-0.02***	-0.02***	-0.017***	-0.017***	-0.017***	-0.017***
BANKS	0.003***	0.003***	-0.007***	-0.007***	-0.011***	-0.011***
VOL	-0.94***	-0.98***	-0.96***	-1.024***	-1.025***	-1.11***
r_t		0.01***		0.012***		0.012***
SKEW		-0.056***		-0.078***		-0.1***
Phase4			-0.007***	-0.007***		
2014					-0.003***	-0.003***
2015					-0.006***	-0.006***
2016					-0.008***	-0.008***
2017					-0.009***	-0.009***
2018					-0.01***	-0.01***
2019					-0.012***	-0.012***
2020					-0.013***	-0.013***
2021					-0.018***	-0.018***
2022					-0.019***	-0.019***
2023					-0.019***	-0.02***
2024					-0.019***	-0.019***
2025					-0.019***	-0.019***
2026					-0.023***	-0.023***
R-adj	50.7	51	59.8	60.1	66.1	66.4

Table 5.1: Table of results obtained from regressions using Models I-VI in Phase 3. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

A relevant result to be seen is that the RATE coefficient has now negative sign, this coefficient is more realistic than the one observed in Phase 2. In fact, considering equal credit, when interest rates grow we have an economic contraction, since the consequence of this increase is the growth in default rates, and consequently the credit spread increases in module. Obviously, since the spread is considered with negative sign, it decreases. This result suggests seeing the spread observed on EUAs market like a credit spread.
Another relevant observation to make concerns the change in sign of BANKS variable happening from Model III, this is due to the introduction of the dummies, since the Phase4 variable is highly correlated with the BANKS variable (-65.02%) and so are the maturity dummies. Finally, the returns and skewness variables, like before, do not have significant impact on the $adjusted-R^2$ of the model, hence can be considered not relevant.

5.2. The Role of Volume

In this paragraph, we present the implementation of Models from I to VI with the introduction of the volume variable and the analysis of the models considering only the observations filtered with the ICE volume filter proxy.

Firstly, the volume variable is introduced. From the results obtained in Phase 2 one can expect that the $adjusted - R^2$ increases when the variable is introduced, especially for the first two models, while should not expect a relevant increase in explanatory power for the last four since we already consider similar information via the dummy variables. Table 5.2 shows the results obtained applying the regression models.

Model	*I*	*II*	*III*	*IV*	*V*	*VI*
Intercept	-0.03***	-0.03***	-0.023***	-0.023***	-0.013***	-0.012***
RATE	-0.021***	-0.021***	-0.019***	-0.019***	-0.018***	-0.018***
BANKS	0.004***	0.004***	-0.003***	-0.003***	-0.008***	-0.008***
VOL	-0.97***	-1.02***	-0.98***	-1.041***	-1.017***	-1.09***
VOLUME	0.41***	0.41***	0.29***	0.29***	0.158***	0.16***
r_t		0.011***		0.012***		0.012***
SKEW		-0.06***		-0.073***		-0.09***
Phase4			-0.005***	-0.005***		
2014					-0.002***	-0.002***
2015					-0.004***	-0.004***
2016					-0.005***	-0.005***
2017					-0.006***	-0.007***
2018					-0.007***	-0.007***
2019					-0.009***	-0.009***
2020					-0.01***	-0.01***
2021					-0.014***	-0.014***
2022					-0.015***	-0.015***
2023					-0.015***	-0.015***
2024					-0.015***	-0.015***
2025					-0.014***	-0.014***
2026					-0.018***	-0.018***
R-adj	60.1	60.3	63.9	64.1	66.9	67.2

Table 5.2: Table of results obtained from regressions using Models I-VI in Phase 3, with the addition of the volume variable. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

As expected, the value of $adjusted - R^2$ increases significantly in Model I and Model II with respect to the same regressions without the volume variable. However, considering the best performing model the increase is less than 1% hence, as for Phase 2, the maturity dummies catch almost the same information, namely the decrease in module of the spread as date of maturity is approached.

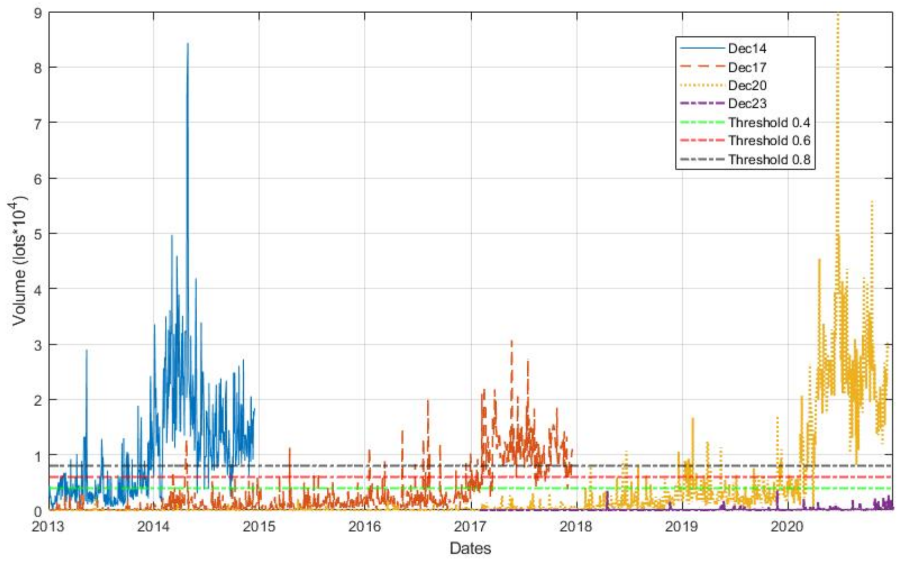


Figure 5.2: Graph of the volumes of futures traded, expressed in 10000 lots. The contracts considered are the ones maturing in: December 2013 (in blue), December 2017 (in orange), December 2020 (in yellow), and December 2023 (in purple). The red dotted line represents the ICE proxy volume threshold, while the black and green lines represent the consistency check thresholds. Data Source: Refinitiv Eikon.

The second study proposed is the analysis of observations which satisfy the designed filter $\frac{Volume}{60} \geq 100$. As for previous case, the analysis is also proposed with a more relaxed condition and a stringent one to have a reliability check, these are done reducing the condition to $\frac{Volume}{40} \geq 100$ or increasing it to $\frac{Volume}{80} \geq 100$ and can be seen in Appendix B. Table 5.3 shows the results obtained using the first presented filter.

Model	*I*	*I*	*II*	*II*	*III*	*III*
Intercept	-0.019***	-0.022***	-0.019***	-0.022***	-0.018***	-0.021***
RATE	-0.016***	-0.019***	-0.016***	-0.019***	-0.02***	-0.017***
BANKS	0.0005	0.003**	0.0004	0.003**	-0.0009	0.0012
VOL	-0.52***	-0.56***	-0.56***	-0.597***	-0.53***	-0.55***
VOLUME		0.104***		0.104***		0.084***
r_t			0.008*	0.008*		
SKEW			-0.052	-0.051		
Phase4					-0.006***	-0.005***
R-adj	47.2	48.5	47.4	48.6	48.5	49.3

Table 5.3: Table of results obtained from the regressions with Models I-III on the filtered dataset. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

In this case the models proposed can stop at the third case, since the introduction of the variables of returns and skewness does not improve the model, hence we can discard

Model IV and Model VI. Moreover, the dummy variables related to the maturity of the contracts increase the multicollinearity index, hence we can stop at the introduction of the dummy of phase that partially substitutes the other dummies, therefore we can ignore Model V.
It is then important to notice that the variables of returns and skewness are statistically not significant at 1% level, so the result coincides with the one obtained for Phase 2 analysis. A particular mention has to be made for the banking variable; in fact, it becomes significant only when the volume variable is introduced. One can expect it to be not significant since it is constructed via a linear interpolation based on annual observations. This is partially confirmed by the high p-value observed in models without volume variable, even if when the latter is introduced the banking variable becomes relevant at least for the first two models.
However, in the final models the variable related to the level of storage can be neglected as its p-value is higher than 1%. This fact confirms that the dummy of phase, together with the volume variable, catches the information relative to the storage, and that the construction of the banking variable might be too imprecise.

5.3. Commodities Analysis

In this paragraph the focus is shifted to the analysis of the spread values using variables related to commodity prices, in particular the regression considers variables related to oil, power and gas prices and their volatilities. Once again, the coal price is not considered, since the different trading days in its market significantly reduce the number of observations that can be used in the regressions. However, now the gas price can be included as its time series starts from 2010.
In Table 5.4 we can see the results of the regressions with and without the use of the filter related to the volume of EUA futures contracts traded.
As for Phase 2 results, the table presents the regressions before and after the removal of variables due to high VIF and high p-values ($\geq 1\%$). The firs clearing is done to delete multicollinearity inside the regression, hence in case of VIF values above 10, one variable at a time is deleted till every VIF is under the threshold, when clearing we prefer to hold inside the regression variables related to prices of commodities since are directly observed on the market, hence the price volatilities are deleted first. The second clearing, related to the p-value, is done to get the best model keeping only significant variables.
As explained, in the table the first columns show the results of the regressions without imposing the filter on volume while the latter ones present the results obtained applying the filter.

Filter	*Case No Volume Filter*		*Case Volume Filter*	
	Before Clearing	After Clearing	Before Clearing	After Clearing
Intercept	-0.036***	-0.021***	-0.027***	-0.018***
RATE	-0.013***	-0.021***	-0.014***	-0.016***
BANKS	0.001***	0.005***	0.001	
VOL	-0.609***	-0.89***	-0.407***	-0.55***
VOLUME	0.41***	0.41***	0.128***	0.102***
Oil Price	0.0002***	-1.00E-04***	9.00E-05*	-4.00E-05**
Oil Volatility	-3.00E-06***		-1.00E-06***	
Power Price	-2.00E-05***	-3.00E-05***	-1.00E-05	
Power Volatility	-2.00E-06***	-3.00E-06***	-9.00E-07**	-2.00E-06***
Gas Price	0.002***	1.00E-04***	0.001***	0.0001**
Gas Volatility	-5.00E-05***		-3.00E-05***	
R-adj	74.2	67.4	52.7	49.7

Table 5.4: Table of results obtained from regressions with commodity models before and after clearing variables due to high multicollinearity or high p-value. On the left side are the regressions obtained considering all the observations, and on the right side are the results obtained considering observations with a volume traded over the threshold of 6000 lots. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

As can be seen from the table the introduction of commodities improves the model, especially for what concerns the regression with all observations, whose $adjusted - R^2$ increases of more or less 7%. However, looking at the same regression applying the volume filter we can notice that the increase in explanatory power is not as big as the previous case. In fact, even if the variables selected in both models are almost the same, the increase in $adjusted - R^2$ is of only 2.5% which shows low dependence from this type of variables.

We can then summarize the results saying that the commodity variables help the explanation of this phenomenon. This is especially true when considering all the observations, while the dependency becomes less relevant when we focus on the observations with more volume, which represent the ones with the most reliable prices.

5.4. Z-Spread Model

In the following paragraph we take a look at the relation between the spread in EUA market and the Z-spread, hence we consider the spread between the risk-free rate and the rate derived from corporate bonds.
For this study the major emitters of GHG are considered. To do so, via Refinitiv Eikon, we obtained the data relative to the Scope 1 emissions, which refer to the emissions caused directly by the companies. To have a double check, via the EUTL we obtained data

relative to verified emissions, allowances allocated, and allowances submitted per year associated to a stationary installation. However, since the only way to recover the firm connected to a specific OHA is by cross-referencing multiple datasets, which sometimes miss firm names, the Scope 1 dataset is held as a reference.
After selecting a dozen firms, we downloaded their coupon bond price histories starting from January 2013. Then, ensuring that for every contract the issue amount is greater than or equal to 500Mil, one can obtain the Z-spread value corresponding to the maturity of each contract, and from them can derive the Z-spread for a specific point in time.
We have selected as time horizon for the Z-spread 2 years since the 3-months spread is less important from the point of view of a factory. Hence, after having obtained the Z-spread at 2 years for all the firms selected, one can compute a simple average to obtain a mean value of it.
In the study two approaches are followed, the first one considering all the selected firms spreads, and the second one using the values of Z-spread of firms with a high difference in free allocated allowances and total emissions, data obtained via EUTL cross tables. This second study is done to consider the Z-spread of firms which need to buy more allowances, said spread seems to have more relevance with respect to the first one. This confirms that the Z-spread generated by the firms in shortage of allowances is more related to the allowances market than the one generated via firms that do not need many contracts.
To be clear on the notation, the spread which considers all contracts is identified as Z_{All} while the one which considers only firms with high difference between allocations and verified emissions is defined as Z_{Power} as the energy sector is the main representative of this class. When no data relative to allocations is available, the firm is included in the Z_{Power} category.
The firms selected are presented in Table 5.5.

Firm	**RIC**	**Delta: Emissions vs Free EUAs**
ArcelorMittal SA	MT.AS	Low
CEZ as	CEZP.PR	High
HeidelbergCement AG	HEIG.DE	Low
Thyssenkrupp AG	TKAG.DE	Low
Eni SpA	ENI.MI	High
Engie SA	ENGIE.PA	High
Veolia Environnement SA	VIE.PA	High
Electricite de France SA	EDF.PA	High
Compagnie de Saint Gobain SA	SGOB.PA	N/A
BASF SE	BASFn.DE	High
Enel SpA	ENEI.MI	High
Fortum Oyj	FORTUM.HE	High
AP Moeller	MAERSKb.CO	N/A

Table 5.5: Table of selected firms used to calculate the Z-spread, specifying if a firm's difference in emissions and free allocations is high, low or not available. As can be seen most of the firms with high delta belong to the energy sector.

Figure 5.3 gives a visual representation of the Z-spread when considering all the firms and when they are filtered.

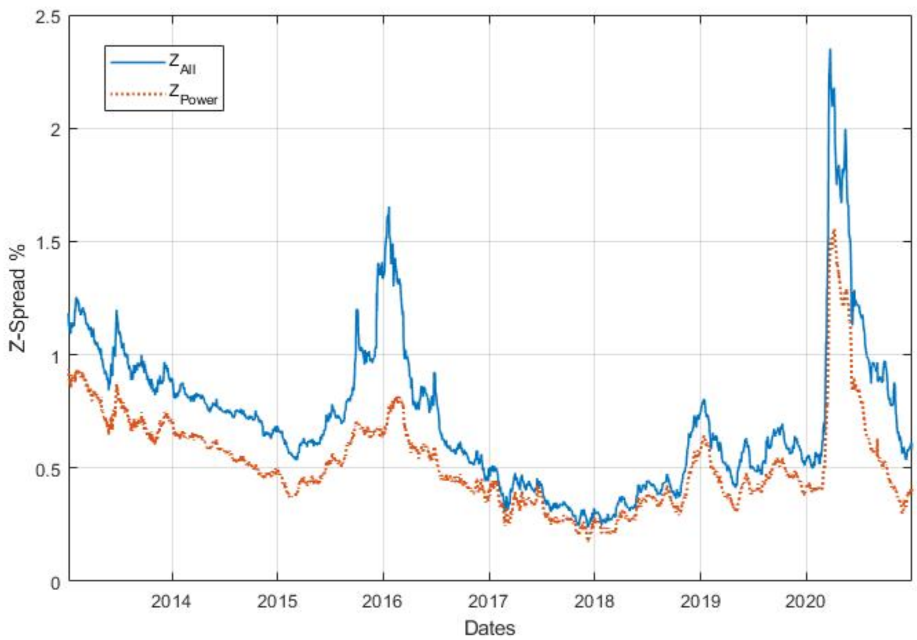


Figure 5.3: Graph of Z-Spreads obtained considering all firms (in blue) and only the ones with high difference between free allocations and total emissions, mainly represented by power plants (in orange). The values are expressed as percentages.

In Table 5.6 we present the results obtained with the use of Model I considering the introduction of Z-spread and the volume variable, as equation (3.12).
In the analysis we consider both the case with all observations and the case with only observations with volume higher than the filter threshold, to see if in both situations the Z-spread variable is relevant.

As can be seen, the introduction of the Z-spread variable improves the regression in both cases. From now on we focus on the results with the Z_{Power} variable, which is more relevant, as confirmed by the values of $adjusted - R^2$ in the table, since the data catches the need of contracts from the firms.
When considering all observations one can notice an improvement on the explanatory power of the model, this increases of about 1.4%, moreover the p-value of the variable is not high at all, meaning that it is significant in the regression. Even when filtering the observations, we can notice an improvement in the value of $adjusted - R^2$ which in this case increases of slightly more than 2%. Once again, the p-value of the variable is low enough to consider it to be relevant in the regression.

Filter	*Case No Volume Filter*		*Case Volume Filter*	
	Z_{Power}	Z_{All}	Z_{Power}	Z_{All}
Intercept	-0.024***	-0.026***	-0.017***	-0.019***
RATE	-0.017***	-0.019***	-0.015***	-0.016***
BANK	0.0016***	0.0026***	0.0003	0.001
VOL	-0.85***	-0.92***	-0.47***	-0.51***
VOLUME	0.41***	0.41***	0.11***	0.10***
Z-Spread	-0.60***	-0.24***	-0.48***	-0.21***
R-adj	61.5	60.6	49.5	49

Table 5.6: Table of results obtained from regressions using the Z-Spread Model. Z_{All} represents the spread derived using all the considered firms, while Z_{Power} considers only firms with a high difference between free allocations and verified emissions, mainly represented by power plants. On the left side, the results with all the observations are presented, while on the right side, the results on the filtered observations are presented. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

5.5. Autoregressive Model

The final analysis proposed is the study of the spread in EUAs market as an autoregressive process. After the addition of the lagged spread inside the regression model we can notice that the value of $adjusted - R^2$ shoots over 75% in both the filtered case and the not filtered one. Table 5.7 shows how the model improves when considering the lagged spread.

Filter	*Case No Volume Filter*	*Case Volume Filter*
Intercept	-0.0014***	-0.005***
RATE	-0.001***	-0.004***
BANK	0.0003**	0.0008
VOL	-0.041***	-0.13***
VOLUME	0.022***	0.029**
$Spread_{t-1}$	0.95***	0.78***
R-adj	96.5	79.8

Table 5.7: Table of results obtained from regressions using Model I with the introduction of the volume variable and the lagged spread. On the left side, the results with all the observations are presented, while on the right side, the results on the filtered observations are presented. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as %.

After this analysis we assumed that the process could be considered autoregressive.
Hence, to have a continuous time series, inspired by Palao and Pardo (2021) we roll the spread, as shown in Figure 5.4. In this process the values of spreads get linked as the contract reaches maturity.

To be clearer, consider the December 2013 contract, the spread derived on that future is considered until maturity is reached. When the last possible spread value is set in the vector, the spread obtained with the future maturing in December 2014 is linked to the time series, and we proceed this way till we cover all Phase 3.
Rolling in this way we can obtain a time series which is continuous and that considers only the liquid trades, so we can expect similar results to the one observed applying the volume threshold.
Once a continuous series is obtained, we can focus on the analysis.

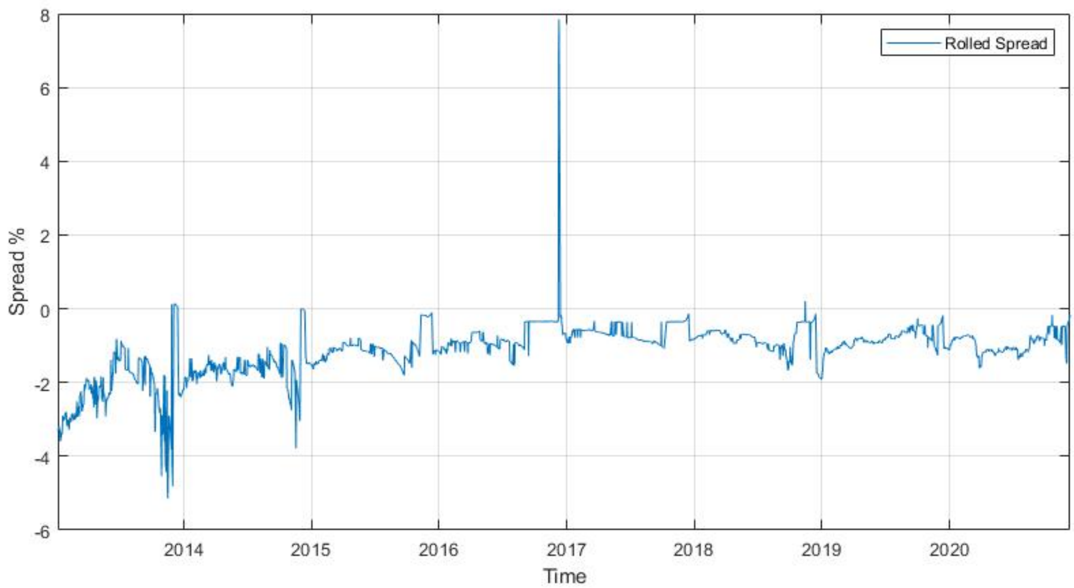
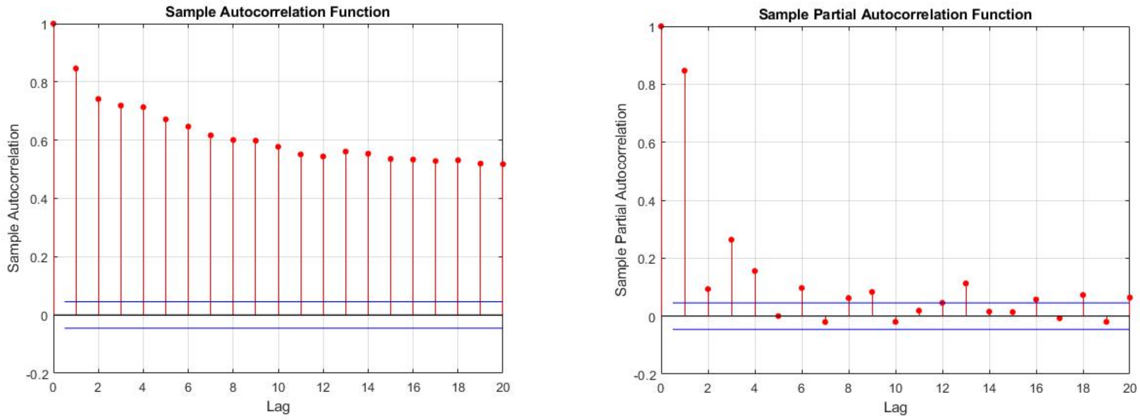


Figure 5.4: Graph of the Rolled Spread in EUAs futures market.

As can be seen, when maturity is approached the spread tends to reduce and fluctuate near the zero value. The only exception is represented by the December 2016 futures, whose spread spikes far above in positive value. This data can be considered as a consequence of the use of EUA daily futures prices as the spot price. In fact, in these dates the spot price is above the futures price, hence these few observations can be treated as outliers.
Moving on to the analysis of the process, as a first step some confirmation of the AR behavior is needed. We decide to use ACF and PACF functions to check the influence of lagged values on the process. Figure 5.5 shows that the only relevant lag to consider is the first one.
After this confirmation, we need to check that the series is stationary, to do so we use an Augmented Dickey–Fuller test, whose result confirms the stationarity of the process as the p-value obtained is lower than 1%.



(a) ACF Function (b) PACF Function

Figure 5.5: Graphs displaying autocorrelation and partial autocorrelation functions of the process. The PACF function indicates that the only relevant lag to consider is the first one.

Now that we have an idea of how the spread evolves, we proceed to its analysis via the use of an OLS regression selecting the variables used in Model I considering also the volume variable, the Z-spread (selecting Z_{Power}) and the lagged process, as in equation (3.15). The first analysis includes the values of spike, it is presented in Table 5.8, and shows that the relevant variables for the model are the 3-months German rate, the variance of spot returns and the lagged values of spread.

Variables	*Coefficient*	*P-Value*
Intercept	-0.004***	3E-7
RATE	-0.0033***	1E-6
BANK	-3E-7	0.58
VOL	-0.19***	0.0002
VOLUME	1E-8	0.18
Z-Spread	-0.11*	0.019
$Spread_{t-1}$	0.72***	0
R-adj	74.9	

Table 5.8: Table of results obtained from the regressions with AR-Model. Both the coefficients and the p-values are reported. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

The results of the regression confirm the AR behaviour of the process but show that the Z-spread and the volume variable, which seemed to be relevant variables to introduce in

the regression, are not significant at 1% value.
To have a double check we propose the same analysis truncating the spike in spread. To do so we consider as limit the 99.9 percentile and delete from the time series the values of spread above the set threshold, a representation of this can be seen in Figure 5.6. The choice of this threshold is done to be as conservative as possible and keep the majority of observations in the time series.

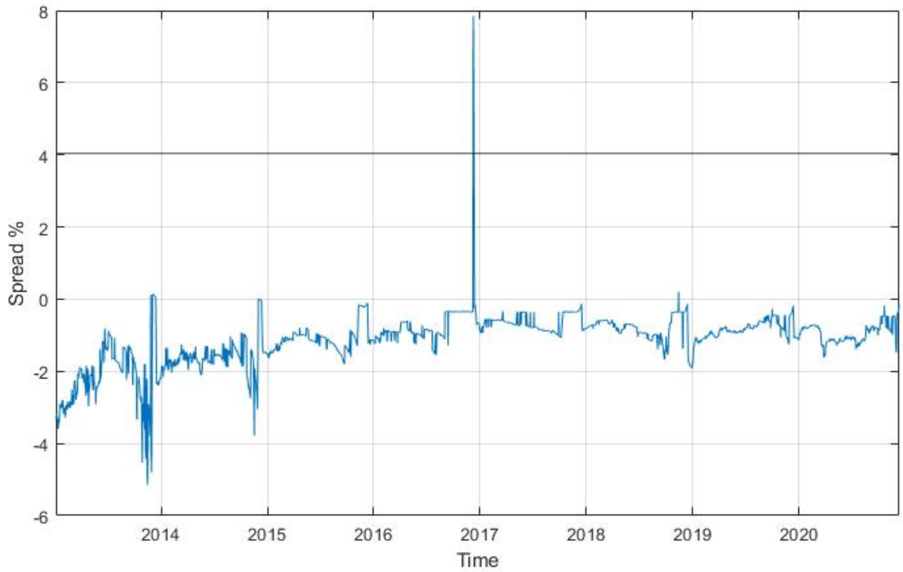


Figure 5.6: Graph of the Rolled spread in EUAs futures market. The black line denotes the 99.9 percentile of the spread values. The observations above the black line are excluded from the time series.

Having deleted the outliers, the regression is proposed one more time, to verify if the previous results were influenced by those extreme values.

Variables	*Coefficient*	*P-Value*
Intercept	-0.004***	9E-9
RATE	-0.0029***	6E-7
BANK	-1E-6*	0.03
VOL	-0.23***	2E-7
VOLUME	1E-8	0.09
Z-Spread	-0.16***	0.0001
$Spread_{t-1}$	0.67***	0
R-adj	78.2	

Table 5.9: Table of results obtained from the regressions with AR-Model on the rolled spread without outliers. In this case Z-Spread considers the Z_{Power} values. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

After the analysis we can assert that the spikes distort in some sense the results of the regression. As it is not realistic to have a positive spread, the results obtained with the removal of the outliers are assumed to be more consistent.
From this study we can deduce that the variables more related to the AR process are the 3-months German rate, the spot returns volatility, and the Z-spread. This means that the volume variable, relevant in previous models, is not considered as a fundamental regressor. The reason behind this might stand in the fact that we do not have multiple values which evidence a different behaviour reaching maturity, but a continuous time series that has always high levels of volume.
To have a double check, the regression has been proposed also with the Z-spread derived from all firms, which of course gave similar results, but with a slightly lower value of $adjusted - R^2$ and a higher p-value of the variable, which still remains under 1%. The results of the regression can be seen in the following table.

Variables	*Coefficient*	*P-Value*
Intercept	-0.0041***	3E-11
RATE	-0.0032***	4E-8
BANK	-1E-6	0.07
VOL	-0.25***	1E-8
VOLUME	1E-8	0.11
Z-Spread	-0.07**	0.0066
$Spread_{t-1}$	0.67***	0
R-adj	78.1	

Table 5.10: Table of results obtained from the regressions with AR-Model on the rolled spread without outliers. In this case Z-Spread considers the Z_{All} values. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

Other checks that are performed see the application of the models after the removal of the tails in the spread distribution, presented in Appendix B. The results of the regression are improving, as the $adjusted - R^2$ increases, and more variables seem to get relevant, as all p-values decrease. Anyway, doing so we risk to no longer have a continuous time series. In fact, cutting the 99.9 percentile was justified as the shooting values were considered outliers and the time series did not get a relevant cut, however truncating at higher percentiles and on both sides of the distribution one might generate too many holes in the distribution. So, we prefer to stick with the results obtained with the 99.9 percentile truncation.

To sum up the results, we deduce that the spread in EUAs market can be interpreted as an AR(1) process, and that the only relevant lag for the event at time t is the one at *t-1*. Moreover, it is proven that the other main drivers of the process are the German short-term rate, the spot returns volatility, and the Z-spread.

5.6. Options Study

At last, we present the study that conducted on option contracts traded on the market. The options on futures are European style options that have contract size of one lot EUA futures, which as previously said have size of 1000 European Union Allowances. Being a European style option, at maturity the options which are in the money are automatically exercised, and the maturity date is defined as the third business day before the underlying futures contract expiry, as specified in ECX (2010).

Following this brief introduction on the options characteristics, we can now move to the conducted analysis. Specifically, after having obtained the price history of put and call options on a certain underlying at set maturity, one can select a strike price to generate a synthetic forward as done in Azzone and Baviera (2021). Using the put-call parity condition one can define the synthetic forward G as contract composed of a long call and a short put on the same strike and at same maturity. From this one can obtain an equation like (5.1).

$$F = \frac{G(K)}{\overline{B}(t_0, T)} + K. \tag{5.1}$$

Where F represents the forward price, $\overline{B}(t_0, T)$ defines the market discount factor, G the synthetic forward and K the strike price. From this, a linear regression approach on G can be followed to obtain the values of $\overline{B}(t_0, T)$. To simplify the procedure, the estimate of these values can be obtained following equation (5.2).

$$\overline{B}(t_0, T) = \frac{\sum_{i=1}^{N}(K_i - \hat{K})(G_i - \hat{G})}{\sum_{i=1}^{N}(K_i - \hat{K})^2}. \tag{5.2}$$

In the equation i rolls on the number of strikes N, $\hat{K}$ represents the mean of the strikes for contracts at a set maturity and $\hat{G}$ the mean of synthetic forwards built with contracts maturing in T at different strikes.

Before applying the equations, one should filter the contracts to avoid the use of penny options, which are options with very low price, and options with wide difference between bid and ask prices.

Looking at the dataset relative to options, which is downloaded from Refinitiv Eikon, one

can see immediately that the volume of options traded and the relative open interest is almost not existing in every contract. This gives a hint that the study conducted might not be reliable.
Despite this, we proceed in the analysis and try to obtain a market discount factor from the options, but the result obtained is not what we expected as all discount factors, at every maturity, seem to be unitary. In fact, low liquidity levels in the market imply that the prices are imposed by the exchange and consequently are not market related.
This study is conducted on ICE options, and as a double check we also try to change exchange and select the EEX dataset. This shift is not helpful as the liquidity on EEX options is even lower than the one on ICE, hence the analysis yields no results.

6 | Conclusions and future developments

In this thesis we derived the main drivers of the behaviour of the spread observed in the EUAs market, addressing the problem when considering every data of the time series and when considering only the observations with high volumes traded.

The first thing we can observe from the analysis is that the shift in phase brought some changes in the market. In fact, considering the analysis on Phase 2 and on Phase 3 one can find major differences, with the most interesting one being the change in sign of the short-term rate variable when considering e.g. Model I. Moreover, from the analysis of the liquid market on the two time windows one can deduce that the short-term rate is not a key factor for Phase 2 while becomes relevant in Phase 3. This entails that the results on Phase 3 can be considered more reliable, as the amount of free allocations drastically diminished.

When considering the whole dataset one can observe that the main drivers of the spread are the short-term rate, the volatility of spot returns, and the reserve of contracts, as pointed out by Trück and Weron (2016) and we proved that the variable of volume can be considered as a relevant piece of information in the model. Additionally, incorporating commodity variables into the model enhances its performance in terms of $adjusted - R^2$. This means that the EUA market spread is influenced by commodities, especially oil, power, and gas prices. Finally, another improvement in the model is given by the Z-spread variable, in fact this variable shows a low p-value in every regression considered, and increases the $adjusted - R^2$. This means that the financing mechanism of factories is a key factor when analysing this phenomenon.

Focusing on the filtered observations dataset, it can be observed that the results differ in certain cases from the analysis of the full dataset. In particular, the variable bond to the level of storage loses relevance and the commodity variables lose a lot of their explanatory power, meaning that the liquid market is more related to the economic situation in the Eurozone and to the volatility of the EUA market. Again, the Z-spread seems to be a key element in the model as its introduction improves the level of $adjusted - R^2$, confirming the statement in the previous paragraph.

We finally propose a model to analyse the rolled spread, in this we consider only one year per contract, meaning we focus once again on the liquid market. The process can be explained as an AR(1) event, adjusted with the contribution of the short-term rate, the volatility in EUA spot market, and the Z-spread. This represents the best model that we obtained in the analysis, as it has the highest $adjusted - R^2$ and confirms the results found before for the liquid market.

From our results we can conclude that, in the liquid market the spread evolves as an autoregressive process, maintaining the dependence from the three variables presented before. This entails that the contango situation is connected with the European economic scenario, the volatility in the spot market, and the financing mechanism of factories, especially the energy sector. These findings support the idea that the spread in EUA market can be proposed as a market spread instead of a convenience yield.

Some final words should be spent on two topics. Firstly, the analysis proved that the use of the volume variable can be considered as a valid substitute of the dummies proposed in Models I-VI. Said variable can be considered more relevant as it is directly observed on the market. Secondly, the rolling procedure needs a final comment. As we roll the spread, we have an idea of how this phenomenon behaves on the short run, meaning we consider one year for each futures contract, but it would be interesting to also analyse its movements on the longer run even if, in this case, the spread would be less market oriented due to the lack of liquidity in those trading dates.

This analysis is proposed also for the time window in Phase 4, but the particular situation induced by the pandemic and the war between Russia and Ukraine makes the results not reliable. So, it is preferred to exclude this phase from the study. Moreover, due to the short time series available results might not be valid.

Further developments of this study can be represented by the implementation of a more precise "volume" filter, used to select the contracts which satisfy ICE criterion. In this way we would have market orientated prices and the analysis on the spreads derived from those observations would be more robust and precise. Moreover, once the economic situation stabilizes, it would be interesting to study the evolution of this spread also in Phase 4 with the use of more firms to derive the Z-spread. Finally, if the market makes it possible, a final development, would be to propose the analysis made on options which would derive the true market discount factor.

Bibliography

(2010). *ICE ECX Contracts: EUAs and CERs User Guide*. ICE Futures Europe, London, UK.

(2022). *EUA Futures Contract Specifications*. ICE ENDEX, London, UK.

Abrell, J. (2021). Database for the European Union Transaction Log. *Database*.

Alberola, E., Chevallier, J., and Chèze, B. (2008). Price drivers and structural breaks in European carbon prices 2005–2007. *Energy policy*, 36(2):787–797.

Azzone, M. and Baviera, R. (2021). Synthetic forwards and cost of funding in the equity derivative market. *Finance Research Letters*, 41:101841.

Baviera, R. and Cassaro, A. (2015). A note on dual-curve construction: Mr. Crab's Bootstrap. *Applied Mathematical Finance*, 22(2):105–132.

Baviera, R., Nassigh, A., and Nastasi, E. (2021). A closed formula for illiquid corporate bonds and an application to the European market. *Journal of International Financial Markets, Institutions and Money*, 71:101283.

Clarkson, P. M., Li, Y., Pinnuck, M., and Richardson, G. D. (2015). The valuation relevance of greenhouse gas emissions under the European Union carbon emissions trading scheme. *European Accounting Review*, 24(3):551–580.

Duffie, D. (2010). *Dynamic asset pricing theory*. Princeton University Press.

Ellerman, A. D., Marcantonini, C., and Zaklan, A. (2016). The European Union emissions trading system: ten years and counting. *Review of Environmental Economics and Policy*, pages 89–107.

ESMA (2021). Preliminary report emission allowances and derivatives thereof. Technical report, ESMA, Paris, France.

ESMA (2022). Final report emission allowances and associated derivatives. Technical report, ESMA, Paris, France.

Frino, A., Kruk, J., and Lepone, A. (2010). Liquidity and transaction costs in the European carbon futures market. *Journal of Derivatives & Hedge Funds*, 16(2):100–115.

Kumar, S. and Trück, S. (2014). Unbiasedness and risk premiums in the Indian currency futures market. *Journal of International Financial Markets, Institutions and Money*, 29:13–32.

Palao, F. and Pardo, Á. (2021). The inconvenience yield of carbon futures. *Energy Economics*, 101:105461.

Quemin, S. and Pahle, M. (2023). Financials threaten to undermine the functioning of emissions markets. *Nature Climate Change*, 13(1):22–31.

Swinkels, L. and Yang, J. (2022). Investing in Carbon Credits. *Available at SSRN 4225486*.

Tilton, J. E., Humphreys, D., and Radetzki, M. (2011). Investor demand and spot commodity prices. *Resources Policy*, 36(3):187–195.

Trück, S. and Weron, R. (2016). Convenience Yields and Risk Premiums in the EU-ETS—Evidence from the Kyoto Commitment Period. *Journal of Futures Markets*, 36(6):587–611.

Zaklan, A., Wachsmuth, J., and Duscha, V. (2021). The EU ETS to 2030 and beyond: adjusting the cap in light of the 1.5° C target and current energy policies. *Climate Policy*, 21(6):778–791.

A | Appendix A

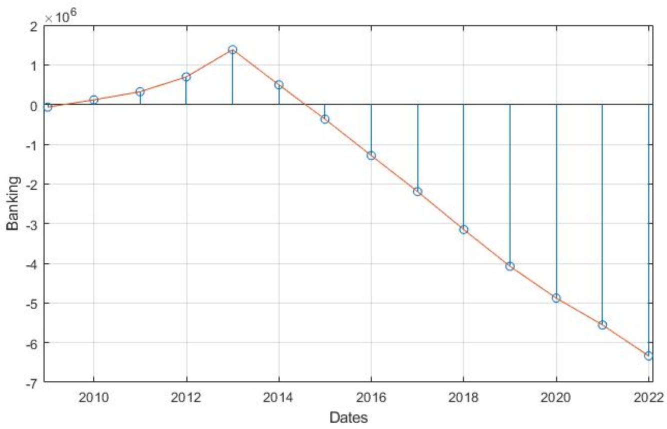


Figure A.1: Graph of the variable Banking in yearly observations and its linear interpolation to obtain daily data. This interpolation is done using free allocations. The unit of measure is k-ton of CO_2 equivalents. Source: European Environment Agency.

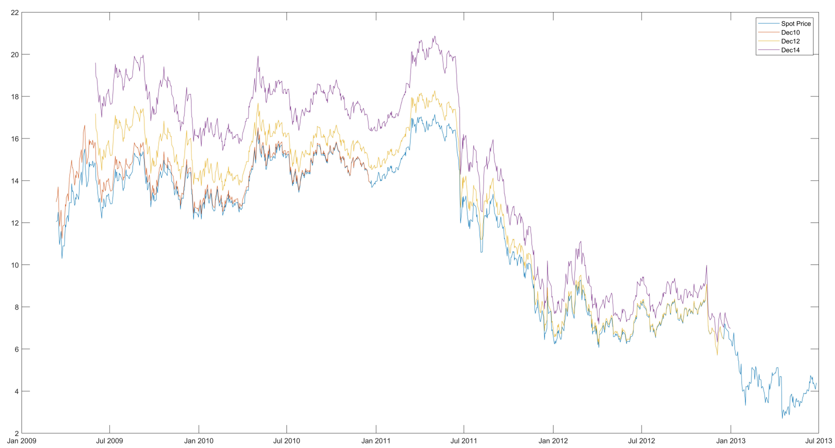


Figure A.2: Graph of spot and futures prices in Phase 2. The spot price (in blue) and futures prices for contracts maturing in December 2010 (orange line), 2012 (yellow line), and 2014 (purple line). Prices are in euros. Data Source: Refinitiv Eikon.

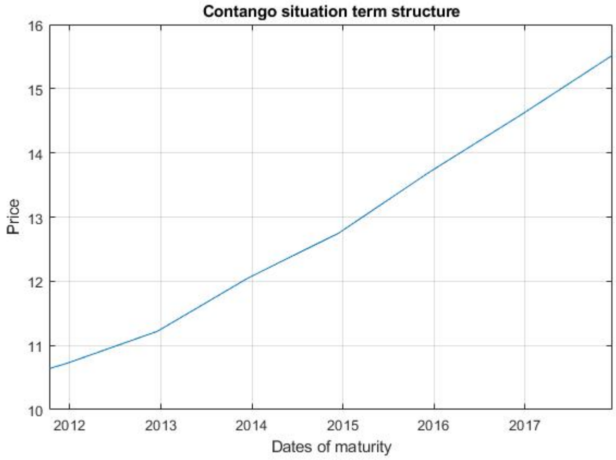


Figure A.3: Example of the contango situation happening in Phase 2, the reference date is October 10, 2011. Prices are in euros. Data Source: Refinitiv Eikon.

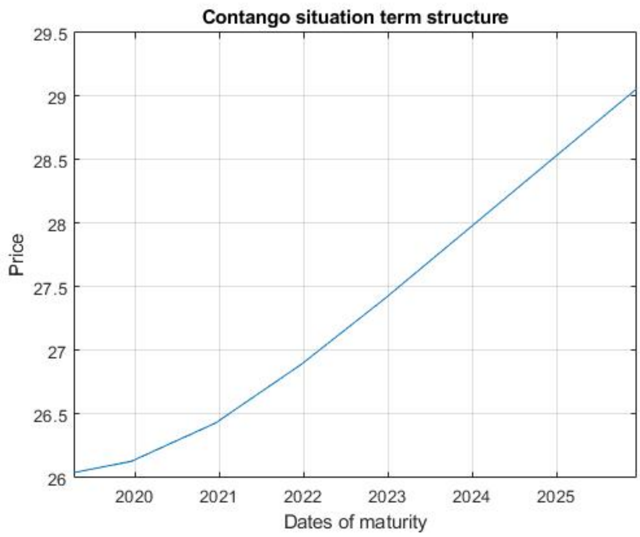


Figure A.4: Example of the contango situation happening in Phase 3, the reference date is April 10, 2019. Prices are in euros. Data Source: Refinitiv Eikon.

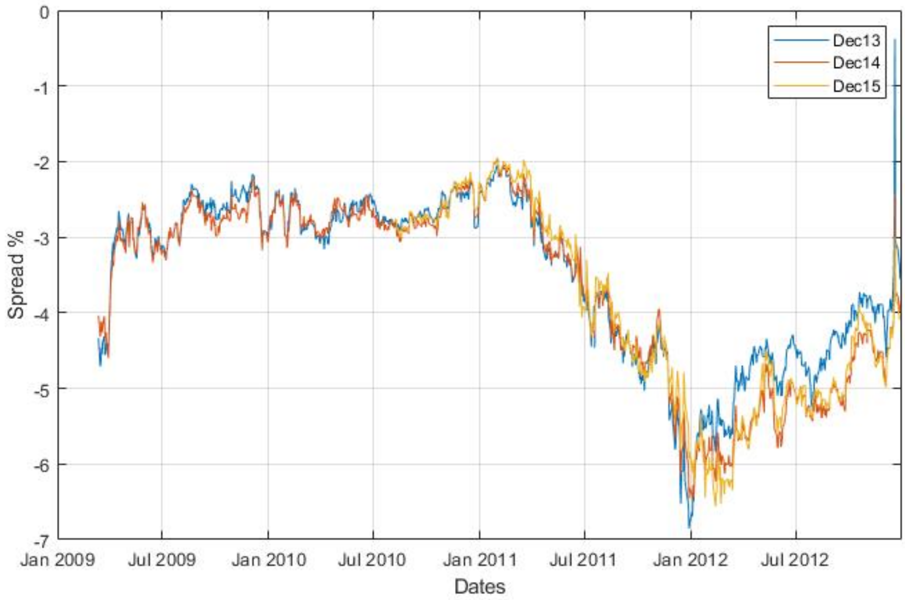


Figure A.5: Graph of spread values, expressed in percentage, for contracts maturing after Phase 2. In particular, the futures contracts considered are the ones maturing in December 2013, December 2014, and December 2015.

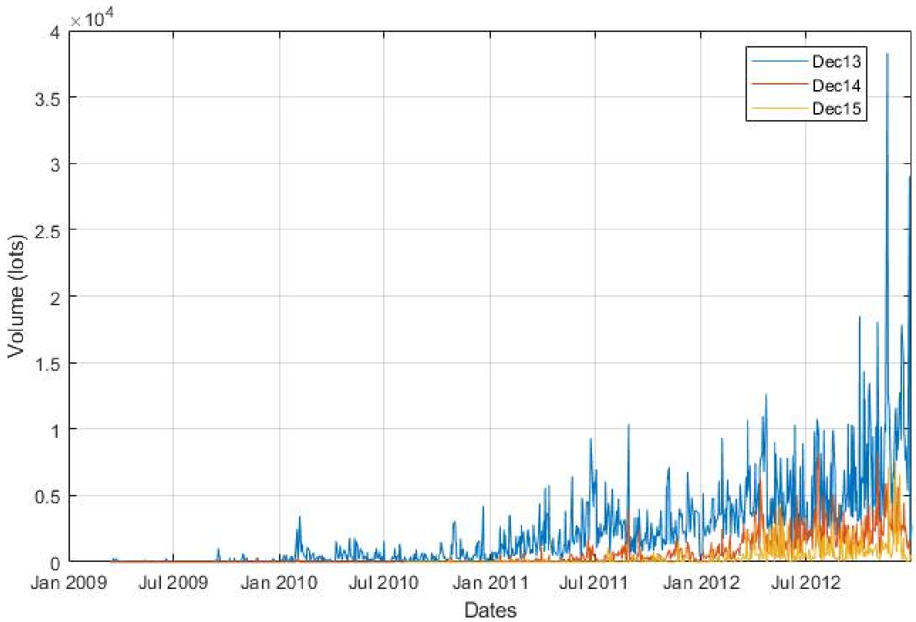


Figure A.6: Graph of traded volumes for futures contracts maturing in December 2013, December 2014, and December 2015, with the values expressed in 10,000 lots. It can be observed that the volume traded reduces as longer maturities are considered. Data Source: Refinitiv Eikon.

Figure A.7: Graph of spread values, expressed in percentage, for contracts maturing after Phase 3. In particular, the futures contracts considered are the ones maturing in December 2021, December 2022, December 2023, and December 2024.

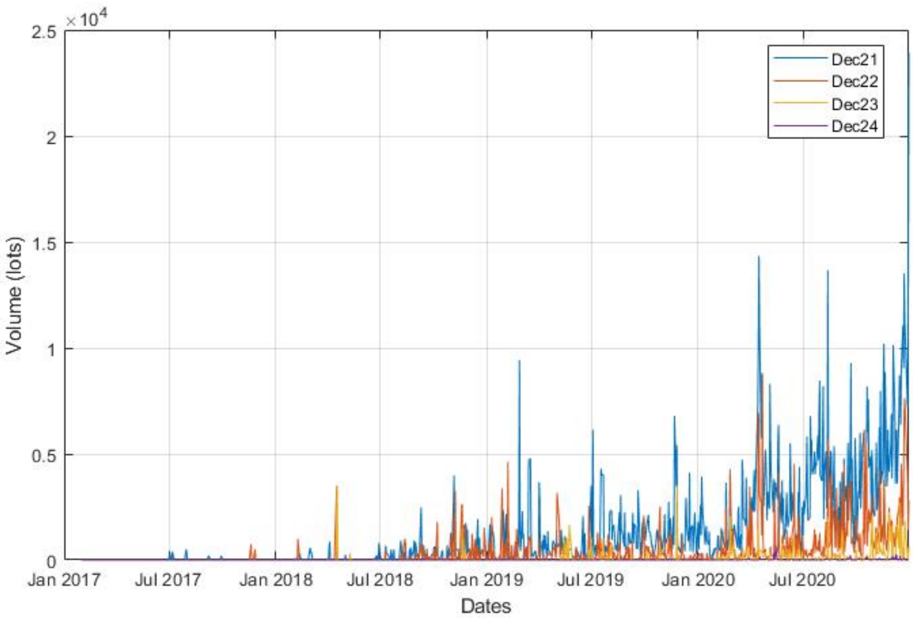


Figure A.8: Graph of traded volumes for futures contracts maturing in December 2021, December 2022, December 2023, and December 2024, with the values expressed in 10,000 lots. It can be observed that the volume traded reduces as longer maturities are considered. Data Source: Refinitiv Eikon.

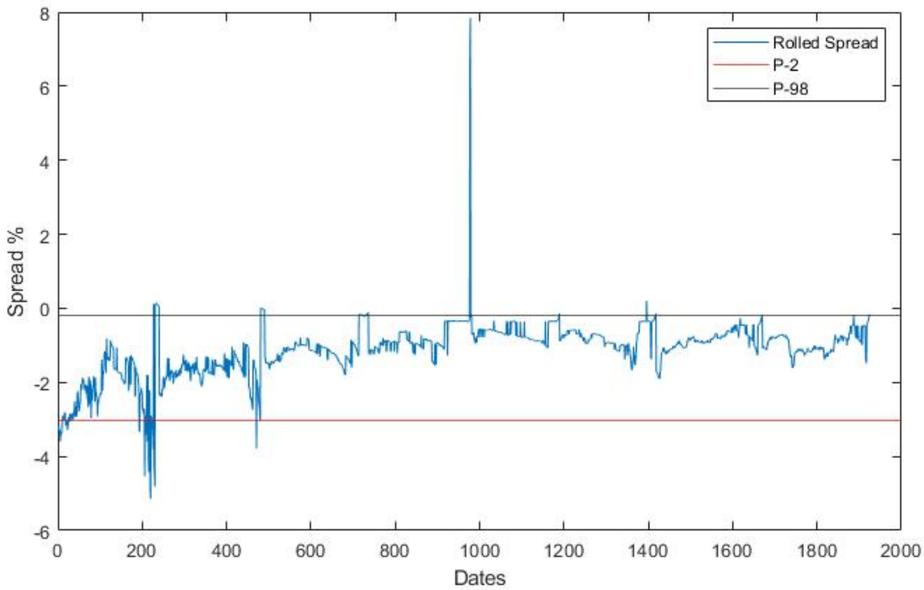


Figure A.9: Graph of the rolled spread in percentage value, with the 2nd percentile and 98th percentile thresholds.

B | Appendix B

In this appendix, the check tables are presented.
For Phase 2 we present the checks related to the filtering methodology. We show the results obtained using regression models on the two other thresholds:

Model	*I*	*I*	*II*	*II*	*III*	*III*
Intercept	-0.008***	-0.01***	-0.007***	-0.01***	-0.009***	-0.009***
RATE	-0.002	-0.002	-0.0018	-0.002	-0.003	-0.003
BANK	-0.03***	-0.03***	-0.03***	-0.03***	-0.021***	-0.021***
VOL	-8.83***	-9.42***	-8.39***	-9.06***	-8.36***	-8.42***
VOLUME		0.3***		0.32***		0.03
r_t			0.028	0.25		
SKEW			0.071	0.22		
Phase3					-0.013***	-0.013***
R-adj	33.3	34.2	34.7	35.7	38.7	38.7

Table B.1: Table of Model I-III regressions with and without the volume variable in case the set threshold is $\frac{Volume}{40} \geq 100$. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

Model	*I*	*I*	*II*	*II*	*III*	*III*
Intercept	-0.008***	-0.008***	-0.008***	-0.008***	-0.009***	-0.008***
RATE	-0.003	-0.003	-0.003	-0.003	-0.004	-0.003
BANK	-0.021***	-0.021***	-0.024***	-0.024***	-0.018***	-0.017***
VOL	-9.14***	-9.07***	-8.5***	-8.43***	-9.02***	-8.76***
VOLUME		-0.03		-0.03		0.13
r_t			0.028	0.28		
SKEW			-0.27	-0.29		
Phase3					-0.01***	-0.01***
R-adj	25.1	25	26.2	26.1	26.6	26.7

Table B.2: Table of Model I-III regressions with and without the volume variable in case the set threshold is $\frac{Volume}{80} \geq 100$. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

As we can see, the $RATE$ variable remains not significant, like the original case. The $VOLUME$ variable seems to be significant when the filter is more relaxed, in the first models, while with the introduction of the dummy its significance reduces confirming the results of the initial filter.

For Phase 3 we first present the results of regressions when considering the other two thresholds and then the checks related to the extension of the observations to Phase 2 data, related to contracts maturing in Phase 3. Finally, we present the results of the regression with lagged spread when we truncate the spread values at its tails.

Model	*I*	*I*	*II*	*II*	*III*	*III*
Intercept	-0.021***	-0.025***	-0.02***	-0.025***	-0.018***	-0.023***
RATE	-0.018***	-0.021***	-0.018***	-0.02***	-0.016***	-0.019***
BANK	0.0014	0.004***	0.001	0.004***	-0.002	0.0019*
VOL	-0.63***	-0.66***	-0.65***	-0.069***	-0.63***	-0.66***
VOLUME		0.18***		0.18***		0.152***
r_t			0.007*	0.009*		
SKEW			-0.05	-0.053		
Phase4					-0.006***	-0.004***
R-adj	50.1	53.8	49.5	53.2	52.2	54.6

Table B.3: Table of Model I-III regressions with and without the volume variable in case the set threshold is $\frac{Volume}{40} \geq 100$. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

Model	*I*	*I*	*II*	*II*	*III*	*III*
Intercept	-0.018***	-0.02***	-0.018***	-0.02***	-0.018***	-0.019***
RATE	-0.016***	-0.017***	-0.016***	-0.017***	-0.016***	-0.017***
BANK	0.0004	0.0017	0.0002	0.001	-0.00028	0.0008
VOL	-0.47***	-0.49***	-0.48***	-0.5***	-0.47***	-0.49***
VOLUME		0.59***		0.05**		0.046**
r_t			0.006	0.007		
SKEW			-0.044	-0.43		
Phase4					-0.006***	-0.005***
R-adj	46.9	47.3	46.1	46.4	47.7	47.9

Table B.4: Table of Model I-III regressions with and without the volume variable in case the set threshold is $\frac{Volume}{80} \geq 100$. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

Again, we have confirmation in results. In fact, in both cases the $BANK$ variable is not significant at 1% level in the last model, while for the remaining variables the sign and significance level coincide.

Now we focus on the check of the whole time series of contracts maturing in early stages of Phase 3. So, for contracts like the 2013 futures we consider data also in days before the beginning of 2013.

Model	*I*	*I*	*II*	*II*	*III*	*III*
Intercept	-0.022***	-0.022***	-0.022***	-0.022***	-0.02***	-0.021***
RATE	-0.013***	-0.012***	-0.013***	-0.012***	-0.015***	-0.013***
BANK	-0.009***	-0.01***	-0.009***	-0.01***	-0.013***	-0.012***
VOL	-1.32***	-1.38***	-1.66***	-1.72***	-1.3***	-1.37***
VOLUME		0.467***		0.464***		0.43***
r_t			0.023***	0.024***		
SKEW			-0.45***	-0.44***		
Phase4					-0.005***	-0.003***
R-adj	39.9	44.3	41	45.4	41.3	44.7

Table B.5: Table of Model I-III regressions, both with and without the volume variable, using all the observations for contracts that mature in Phase 3. This means that data observed in Phase 2 for contracts that mature in Phase 3 are also considered. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

As can be seen, the sign and significance level of the variables are coherent with the ones obtained considering only Phase 3 observations. The only noticeable change stands in the reduction of the $adjusted - R^2$.

Proceeding with the checks, we filter the dataset:

Model	*I*	*I*	*II*	*II*	*III*	*III*
Intercept	-0.025***	-0.03***	-0.025***	-0.03***	-0.025***	-0.03***
RATE	-0.025***	-0.027***	-0.025***	-0.027***	-0.024***	-0.027***
BANK	0.005***	0.008***	0.005***	0.009***	0.005***	0.008***
VOL	-0.55***	-0.65***	-0.67***	-0.77***	-0.55***	-0.65***
VOLUME		0.2***		0.21***		0.2***
r_t			0.014**	0.015***		
SKEW			-0.13*	-0.14**		
Phase4					-0.0004	-0.0001
R-adj	47.8	51.1	48.1	51.4	47.8	51.1

Table B.6: Table of Model I-III regressions, both with and without the volume variable, using all the observations for contracts that mature in Phase 3. In this case we apply the ICE proxy filter. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

In this case the results are coherent only for some variables. In fact, $BANK$ is expected to be not significant, while it stays relevant in this case. Moreover, the variables of returns

and skewness have more importance in this regression. This fact derives from the influence of previous data. however, since the increase in $adjusted - R^2$ is very small, we can say that the introduction of those variables is not significant.
The last observation to make is on the phase dummy, in fact it seems to lose efficiency when these data are introduced. This change is justified by the fact that now we are considering three different phases, since Phase 2 data are introduced, and we use the dummy to get the spread of contracts maturing in Phase 4. Moreover, as we said, the volume variable catches similar information, hence might works as its substitute.

Moving to commodities regressions:

Filter	*Case No Volume Filter*		*Case Volume Filter*	
	Before Clearing	After Clearing	Before Clearing	After Clearing
Intercept	-0.019***	-0.0007	-0.021***	-0.017***
RATE	-0.001***	-0.004***	-0.018***	-0.022***
BANK	-0.001***	-0.005***	0.006***	0.006***
VOL	-0.76***	-1***	-0.29***	-0.45***
VOLUME	0.4***	0.43***	0.17***	0.19***
Oil Price	0.0004***	-4.00E-04***	2.00E-04***	-2.00E-04***
Oil Volatility	-8.00E-06***		-3.00E-06***	
Power Price	-3.00E-05***	-3.00E-05***	-2.00E-05	
Power Volatility	-1.00E-06***	-2.00E-06***	-1.00E-06***	-1.00E-06***
Gas Price	0.0003***	7.00E-04***	-0.0004***	-3.00E-04***
Gas Volatility	1.00E-07	-9.00E-06***	2.00E-05***	2.00E-05***
R-adj	68.1	60.2	58.8	55.9

Table B.7: Table of results obtained from regressions with the Commodity Model before and after clearing variables due to high multicollinearity or high p-value. On the left side are the regressions obtained considering all the observations, and on the right side are the results obtained considering observations with a volume traded over the threshold of 6000 lots. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

In this case, we find consistent results. We can see that the introduction of the commodity variables increases the $adjusted - R^2$ especially in the non-filtered case, which is the same result obtained before.
In the filtered case we can see an increase of $adjusted - R^2$ slightly more relevant, but the results confirm our analysis. This improvement stands in the fact that the banking variable remain significant at 1% level and refines the regression.
The last thing to notice stands in the sign of $GasPrice$ which changes from the original version of the model, this is due to the correlation with its volatility. The remaining variable are selected again using VIF analysis and a p-value deletion criterion, so we

can say that there is not high multicollinearity in the model and the variables are all significant.

The last check shows the analysis of the rolled spread when we delete the tails at 2% and 98%. This table shows the analysis using the Z_{Power} variable.

Variables	*Coefficient*	*P-Value*
Intercept	-0.0046***	9E-24
RATE	-0.0036***	3E-17
BANK	-1E-6**	0.0011
VOL	-0.24***	3E-13
VOLUME	8E-9	0.14
Z-Spread	-0.17***	1E-8
$Spread_{t-1}$	0.59***	0
R-adj	84.4	

Table B.8: Table of results obtained from the regressions with AR-Model on the rolled spread without outliers at 2 and 98 percentiles. In this case Z-Spread considers the Z_{Power} values. The asterisks indicate significance of the variable at the 0.1% (***), 1% (**) and 5% (*) level of significance. The R-adj is expressed as a %.

List of Figures

List of Tables

Acknowledgements

Vorrei ringraziare in primo luogo il Prof.Baviera e Pietro Manzoni, che si sono sempre dimostrati disponibili e soprattutto competenti. I vari momenti di confronto sono sempre risultati stimolanti e molto utili, non solo al fine di produrre la tesi.

Un grazie speciale va alla mia famiglia, che mi ha sempre supportato e sopportato, specie nei periodi di sessione in cui ero particolarmente distaccato e nervoso.
Tra le persone speciali che mi hanno accompagnato non posso non citare Rita, Fede, Rick, Fra e tutti gli amici e colleghi universitari, con cui ho condiviso momenti di gioia e di tristezza per esami e studio, e con cui condivido anche bellissimi ricordi legati alle feste e alle vacanze passate insieme.
Ringrazio anche gli amici e colleghi di UniCredit, che mi hanno sempre aiutato e sono stati molto comprensivi durante il mio periodo di stage.
Infine, vorrei ringraziare gli amici di Cardano, Ele e Bea su tutti mi avete aiutato un sacco a uscire dalla mia bolla e non potrò mai ringraziarvi abbastanza.

Grazie per aver contribuito a farmi essere chi sono oggi.