Emission Allowances in the European Union Emissions Trading System

PART I:

A Time Series Analysis of Emission Allowances in the European Union Emissions Trading System

(Submitted to Università Commerciale Luigi Bocconi in April 2020)

PART II:

Emission Allowances And Natural Gas: A Cointegration Analysis

(Submitted to Stockholm School of Economics in June 2020)

PART I

A Time Series Analysis of Emission Allowances in the European Union Emissions Trading System

Franziska Manke

April 12, 2020

Abstract

This thesis analyses the short term behavior of daily emission allowance (EUA) log returns with a focus on volatility dynamics in the recent market environment. In this thesis, I present a historical overview of the European Union Emission Trading System (EU ETS), analyze the stylized facts of the time series, employ appropriate time series models, and assess model in-sample and out-of-sample performance. Due to the existence of leptokurtosis and volatility clustering in the time series, I implement three GARCH models. In addition to a simple GARCH model, I analyze a GJR-GARCH model and an EGARCH model to assess the existence of a leverage component. The performance of the three models is examined and benchmarked against a naive model, which is not incorporating any conditional variance modeling. I examine the models' performance by conducting in-sample goodness of fit and out-of-sample forecasting analysis. The findings strongly support the appropriateness of models capturing leptokurtosis and volatility clustering in the time series while not unambiguously confirming the suitability of models incorporating a leverage effect.

Contents

1 Introduction

In a world characterized by an increasing focus on climate change and its consequences, the search for a sustainable solution has been intensifying. National governments, as well as the European Union (EU), have implemented an array of regulations and subsidies to incentivize green investments and reduce greenhouse gas (GHG) emissions. One of the key initiatives at the supranational level has been the EU's implementation of the EU Emission Trading System (EU ETS). While the program has suffered from a structural oversupply for an extended period, the EU has implemented various regulatory changes addressing the issue.[1] At the same time, the EU has emphasized its commitment to creating a well-functioning carbon market with the purpose of reducing greenhouse gas emissions in a cost-effective and economically efficient manner (European Union, 2003). Current European Union Allowance (EUA) prices are at a historically high level, and the latest round of regulatory changes starting from 2020 aims to set up the market to function appropriately and incentivize innovation and competition.[2]

In light of the recent EUA market developments and a supportive political backdrop driven by increased environmental awareness, the EUA market will probably play a more prominent role than it has done for the past decade. With increasing importance may come an increasing need for hedgers, risk managers, and speculators to understand the time series behavior of EUAs in order to make appropriate decisions. Therefore, this thesis provides (i) a brief overview of the EU ETS and its historical market developments and regulatory changes; (ii) an updated analysis of EUA log return behavior based on the most recent market data in the third trading period of the EU ETS;[3] (iii) and compares the forecasting accuracy of the time series models. Three different GARCH models are recursively estimated and analyzed to address leptokurtosis, volatility clustering, and leverage effects in the EUA log return series. The models are benchmarked against a

[1] The main regulatory changes that addressed the oversupply are the Commission Regulation (EU) No 176/2014 of 25 February 2014 and the Decision (EU) 2015/1814 of the European Parliament and of the Council of 6 October 2015.

[2] The latest revision of the EU ETS was established by Directive (EU) 2018/410 of the European Parliament and of the Council of 14 March 2018 amending Directive 2003/87/EC to enhance cost-effective emission reductions and low-carbon investments, and Decision (EU) 2015/1814.

[3] The considered time period in this thesis spans from 2016 to 2019 while the full third trading period of the EU ETS started in 2013 and will finish at the end of 2020.

naive model, not incorporating conditional volatility features.

1.1 Overview of the EU Emission Trading System

In 1997 the Kyoto Protocol was adopted, and the EU committed to reducing its greenhouse gas emissions by 8% compared to the 1990 level by the first commitment period from 2018 to 2012.[4] In order to meet its target, the EU adopted Directive 2003/87/EC in 2003, setting out the framework for the EU ETS, which ultimately went into force in 2005 and thereby created a new market. The EU ETS is still the primary tool of the EU's policy to reduce greenhouse gas emissions in line with its climate and energy targets for 2020, 2030, and its long term climate strategy. Alongside other goals in relation to energy efficiency and the share of renewable energy consumption, the EU has set out to cut its greenhouse gas emission by 20% compared to 1990 levels by 2020 and by 40% compared to 1990 levels by 2030.[5,6]

The EU ETS is a cap and trade system, which caps the EU-wide total amount of emissions and allows trading emission allowances in the secondary market in order for total emissions to be reduced in the most cost-effective and economically efficient manner (European Union, 2003). The emission allowances or EUAs give the right to emit greenhouse gas emissions equivalent to the global warming potential of one tonne of CO_2 equivalent. Each participant must return the number of allowances for each tonne of CO_2 equivalent the company has produced that year. This goal may be achieved by reducing emissions itself or purchasing more allowances to meet the requirements. In cases of non-compliance, a penalty per tonne has to be paid by the participant. The level of the cap, and hence the total amount of allowances available to participants in the

[4]The Kyoto Protocol to the United Nations Framework Convention on Climate Change (UNFCCC) was adopted on December 10, 1997. The European Community signed the Kyoto Protocol in April 1998 and it was ultimately ratified by the European Union and its member states in May 2002.

[5]The 2020 climate and energy package's targets were set in 2007 at the European Council and a set of legislation was enacted in 2009. The three main 2020 targets are: reducing greenhouse gas emission by at least 20% below 1990 levels, increasing the share of energy consumption from renewable sources to 20%, and improving energy efficiency to reduce primary energy by 20% compared to forecasted levels.

[6]The 2030 climate and energy framework was adopted in 2014 at the European Council and targets revised upwards at the 2018 Katowice UNFCCC Conference. The three main 2030 targets are: reducing greenhouse gas emission by at least 40% below 1990 levels, increasing the share of energy consumption from renewable sources to 32%, and improving energy efficiency to reduce primary energy by 32.5% compared to forecasted levels.

system, is reducing every year by 1.74% during the third phase of the EU ETS (European Union, 2009). The framework operates in 31 countries of the European Economic Area (EEA) and covers approximately 11,000 installations as well as 500 aircraft operators.[7] The EU ETS coverage is equivalent to approximately 39% of the total EU's greenhouse gas emissions.[7] The majority of emissions covered under the EU ETS are produced by electricity and heat production, accounting for about 54% of 2018 EU ETS covered emissions.[7]

1.2 Historical regulatory and market developments

In the pilot phase from 2005 to 2007, the allowance cap was set at the national level via National Allocation Plans (NAPs), and the majority of allowances were given to affected business at no charge. Power generators and energy-intensive industries were the only covered industries within the pilot phase. The issuance of allowances was excessive in the first trading period driven by the absence of reliable emissions data and by the existence of an inherent agency problem. The system was inherently incentivizing national governments to give out more rather than fewer allowances to their respective country's industries to ensure the competitiveness of those industries. Prices collapsed early 2006 following the dissemination of emissions data evidencing the oversupply of allowances by the NAPs. In 2007 prices fell to zero due to the oversupply and the system's prohibition to use phase one certificates for compliance in phase two.

The second phase of the EU ETS lasted from 2008 to 2012. In response to market developments during the pilot phase, the NAPs issued 6.5% less allowance.[8] In contrast, the share of freely allocated allowances was still high, with approximately 90% in 2012.[8] In accordance with the Kyoto Protocol, participants in the system were able to use International Credits under the Clean Development Mechanism (CDM) and Joint Implementation (JI) program in order to meet part of their requirements under the EU ETS.[9] The amount of International Credits a participant was able to use was determined by the respec-

[7]As presented in the report from the European Commission to the European Parliament and the Council on the functioning of the European carbon market. COM/2019/557 final/2.

[8]Summary of phase one and two policies provided by the European Commission on https://ec.europa.eu/clima/policies/ets/pre2013_en

[9]Clean Development Mechanism and Joint Implementation program are set out in Article 6 and 12, respectively, of the Kyoto Protocol to the UNFCCC on December 10, 1997.

tive NAP. The CDM allows an industrialized country to invest in an emissions reduction project in developing countries as an alternative to investing in emission reduction in their own country and thereby create certified emission reduction (CER) credits. Similarly, the JI program allows countries to invest in projects in other industrialized countries and thus create emission reduction units (ERUs). The global financial crisis induced a severe economic downturn that resulted in lower emissions and hence lower demand for allowances. The fall in demand is well described by the significant decrease in verified emissions recorded in 2009, as illustrated in Figure 1.[10] While the negative effects on the demand side were quick to realize, the regulatory and rules-driven supply side was not able to adjust, causing a surplus of allowances. This structural imbalance was carried over well into the third trading period. Figure 2 provides an illustration of the development of the accumulated surplus across the second and third periods showing the vast majority of the surplus was created in the second period of the EU ETS.[10]

The third phase of the ETS started in 2013 and will finish at the end of 2020. With the start of the new trading period, major reform came into effect, mainly set out in Directive 2009/29/EC (European Union, 2009). A single EU-wide cap was introduced compared to national caps in the prior two phases. In addition, the cap is reducing every year by 1.74% from a 2013 level of 2.1bn allowances, the level equivalent to the average allowances over the 2008-2012 period. Compared to the first two phases, the amount of freely allocated allowances was significantly reduced, and a harmonized auctioning process was established. The main auctioning market place of the primary market is the EEX in Leipzig, Germany. The auctioning of allowances is the default method of allocating EUAs in the third phase. However, over the full trading period, from 2013 to 2020, an approximate amount of 43% of allowances are still freely distributed.[11] The policies around free allocation are organized by industry and designed to avoid carbon leakage.[12]

[10]The data sources for Figure 1 and 2 are the reports from the European Commission to the European Parliament and the Council on the state of the European carbon market in 2012 (COM(2012) 652 final), the report on the functioning of the European carbon market from 2015 (COM(2015) 576 final), and the report on the functioning of the European carbon market from 2019 (COM/2019/557 final/2).

[11]As presented in the report from the European Commission to the European Parliament and the Council on the functioning of the European carbon market. COM/2019/557 final/2.

[12]Carbon leakage refers to the situation in which businesses make a cost based decision and decide to re-located their installations to international locations not covered by the EU ETS and with lower emissions standards and constraints.

Industries deemed to be significantly exposed to carbon leakage receive a higher amount of freely allocated EUAs in order to ensure the international competitiveness of these industries under the EU ETS. The emissions price continued to suffer from the structural oversupply established during the second phase, as can be observed in Figure 4. More details regarding the oversupply and the interventions taken, namely back loading and the Markets Stability Reserve (MSR), are provided in subsection 1.3.

The fourth trading period will commence in 2021 and last until 2030. The main changes include an accelerated decrease in the emissions cap at 2.2% compared to 1.74% applied during the third phase. The oversupply measures mentioned in the previous paragraph and outlined in more detail in subsection 1.3 will take effect in the fourth trading period. The carbon leakage list has been updated to improve the precision of the targeted goal. In addition to amendments to the existing rules, the fourth phase will see the establishment of new features, including the Modernisation Fund and the Innovation Fund, to help affected sectors fund the investments required to transition to a low-carbon economy. The majority of legislation has been entered into force in April of 2018 under Directive (EU) 2018/410 European Union (2018).

1.3 Structural oversupply and regulatory interventions

In the second period of the EU ETS, the emissions market was characterized by a structural imbalance in demand and supply. Supplied emissions were significantly outpacing verified emissions. The accumulated surplus reflected in Figure 2 shows the majority of the surplus was create in the second phase of the EU ETS. The imbalance can be largely attributed to lower than expected emissions driven by the economic downturn induced by the financial crisis and higher than expected supply driven by a large inflow of International Credits eligible for use in the EU ETS. The eligible International Credits under the Kyoto protocol refer to credits created through the CDM and JI program. In the second phase, the quantity of International Credits used by participants amounted to 1.06bn credits.[13] This amounts to more than 50% of the surplus created in the second phase. The expected amount of International Credits used in the third phase amounts to

[13]Quantification of used International Credits in the second trading period provided by the European Commission on https://ec.europa.eu/clima/policies/ets/credits_en

600m, i.e., a significant decrease compared to the second phase. On the demand side, the reduction of emissions, particularly in 2009, see Figure 1, are highlighting the link of emissions to economic activity. At the same time, the share of electricity generated from renewable sources steadily increased from 17% in 2008 to 32% in 2018, as illustrated in Figure 5.[14] In summary, the combination of an inadequate and immature framework and the coinciding economic environment were critical drivers for the structural imbalance.

With the apparent inadequacy of the existing system, the EU took action in 2014 (European Union, 2014) and announced to postpone the auctioning of 900mt of allowances over the 2014 to 2016 period to the 2019 to 2020 period. This short term measure is referred to as back loading aimed at cutting the supply side to balance demand and supply. The measure succeeded in the sense that during the years 2014-2016 back loading was applied, the surplus shrank. However, structurally there had been no change, and the emissions price remained in a single-digit price range until 2018. To address the issue with a more long term solution, the EU announced the introduction of a Market Stability Reserve (MSR) (European Union, 2015), aiming to create a rules-based mechanism that dynamically adjusts the supply side in reaction to demand developments. The MSR began operating in 2019. Back loaded allowances were put in the MSR rather than being auctioned in 2019/2020. If circulating allowances exceed 833mt as published in an annual report from the Commission, allowances will be placed in the MSR. In contrast, if published figures are below 833mt, allowances are released from the MSR. Throughout 2019 to 2023, in cases the allowances in circulation exceed 833mt, 24% of the allowance surplus is placed in the reserve rather than auctioned in each year. From 2023 onward, an additional mechanism will apply, canceling all previous allowances above the previous year's auction volume. With this comprehensive set of changes and in combination with the accelerated cap reduction by 2.2% from the start of the fourth trading period, the EU has set out a more robust framework that will show its effectiveness in the coming years.

[14] Data reported by Eurostat on https://ec.europa.eu/eurostat/web/products-datasets/-/sdg_07_40

1.4 Price determinants of emission allowances

As Benz and Trück (2009), Christiansen et al. (2005), and others outline the main categories of emission price drivers are the regulatory framework and fundamentals. Having outlined the regulatory framework and historical developments in the previous subsections, it is clear that the long term demand and supply, but particularly the supply, are heavily impacted by the framework mechanics. Similarly, fundamental factors impacting the discharge of emissions are influencing the demand-supply dynamics. The fundamental factors impacting the production level of CO_2 are multifaceted and inter-related due to the implicit connection of the emissions market to other markets, most notably the energy markets. The interconnection extends to both the fuel input, such as the oil and gas markets, but also the output in relation to the electricity market. In addition, weather related increases in production requirements may occur, but also much broader global economic growth is impacting production levels. These relationships are discussed by various authors including Hammoudeh et al. (2014), Aatola et al. (2013), Chevallier (2011), and Mansanet-Bataller et al. (2007). While both regulatory factors and fundamental factors can shift more broadly over time, they can similarly impact short term price and particularly volatility dynamics in the EUA market. On the regulatory side, short term impacts may arise from EU decisions, announcements, and other publications such as the carbon market report containing new information the market has not yet incorporated. While regulatory decisions are discussed thoroughly in advance, small unexpected changes in the outcome of a decision can equally lead to short term effects. In addition to new information, the regulatory framework can also implicitly lead to varying return and volatility dynamics if the mechanics are more or less effective than anticipated by market participants, particularly in the context of rapidly changing fundamental factors. As highlighted by the very much present market volatility due to the coronavirus, changes in economic outlook can change significantly in a short time frame and have an immediate knock on effect on energy and emission prices. Short term volatility of a different type can also arise due to weather related events. Such events can be cold snaps or generally cold winters increasing demand for heating and hence electricity and gas. Consequently, the increasing production of CO_2 requires companies

to purchase more EUAs in the market. With an ever increasing amount of electricity being generated by renewable sources, as previously discussed and illustrated in Figure 5, one may suspect that the weather plays an increasing role in the demand for EUAs. In prolonged, very windy, and sunny conditions, the electricity production mix of a country, can substantially shift towards renewable sources and crowd out fossil fuel based electricity production given their higher break-even production points. The immediate impact in the electricity market is driven by the uniqueness of it with its constant physical requirement of supply and demand balance. In an environment with a high number of periods with extreme weather and the increasing share of renewable sources could lead to the conclusion that the emissions price will experience more periods of higher volatility.
In addition, to events outside of market participants' control, the price of EUAs may also be impacted by the direct decisions of emitters on their hedging programs, fuel switching decisions, and more broadly renewable investment decisions. Again both short and long term dynamics may be impacted directly by the aforementioned decisions. As a number of emitters, particularly large utilities, are responsible for a large share of emissions covered under the EU ETS, the decision of one may have a significant impact on overall prices. The decision information is asymmetric in nature, and hence other market participants would not be aware before the decision is announced or action is taken, consequently potentially leading to short term shocks in return and volatility. Overall the emissions allowance price is impacted by a variety of inter-correlated and uncorrelated factors that can have both long term as well as short term implications.

This thesis analyses the short term implications for the log return behavior of EUAs with a focus on the volatility dynamics in the recent market environment. The performance of three different GARCH models is examined and benchmarked against a t-distribution (T-DIST). All three GARCH models should outperform the simple t-distribution if volatility clustering is a pronounced feature in the data. Both GJR-GARCH and EGARCH should outperform the simple GARCH model if there is a leverage effect in the volatility process. EGARCH should outperform if non-negativity constraints in the parameter estimation for GARCH and GJR-GARCH are active and if the leverage effect is in fact reversed and positive shocks are more pronounced. The results of this thesis indicate volatility clustering and leptokurtosis are prominent features in the data, while asymmetry is not unequivocally supported. The results can be used in order to choose an appropriate model to forecast short term volatility and returns in the context of the most recent market environment, capturing underlying regulatory developments and fundamentals. The remainder of the thesis provides a review of the existing literature, discusses methodologies used in order to complete the aforementioned analysis, and presents the empirical results.

2 Literature review

The literature on emission allowances investigating the return dynamics from an econometric and risk management angle has seen rapid advancement since the inception of the market in the mid 2000s. While the market matured and regulatory changes have been implemented, underlying market dynamics have changed as well. Particularly in the first trading period, the findings are highly dependent on the time period, which is analyzed as can be easily deduced from the erratic price movements in the first phase of the EU ETS, depicted in Figure 4. In the period shortly after the inception of the emission allowances market in the EU ETS, both empirical studies, as well as theoretical models, were developed to capture EUA price dynamics and market mechanics.

In a theoretical approach Fehr and Hinz (2006) describe a model focusing on the energy sector's agents decisions in regards to short term abatement levels, i.e., fuel switching, based on the cheapest available options. In contrast, Seifert et al. (2008) develop a stochastic equilibrium model in which each affected company decides to spend the optimal amount of money on lowering emissions with the total expected amount of emissions as a key decision driver. The authors find that a CO_2 price process should not necessarily be characterized by seasonal patterns but should possess the martingale property and exhibit time and a price dependent volatility structure. Chesney and Taschini (2012) construct an endogenous model generating the price dynamics of emission permits under asymmetric information, allowing inter-temporal banking and borrowing. The equilibrium permit price is determined by optimizing each firm's decision and imposing the market clearing condition.

In an early study Christiansen et al. (2005), identify key drivers impacting EUA market prices: regulatory and policy issues, market fundamentals, and technical analysis. The authors highlight the importance of the regulatory and policy issues during the first trading period while expecting the increase of the significance of fundamental factors going forward. Fundamental factors are including the CO_2 production, which itself is driven by factors including weather, fuel prices, and economic growth as well as the emission cap and supply of credits from CDM projects. Benz and Trück (2006) investigate stylized facts of the EU ETS and draw from market experience in the SO_2 market in the US. Similarly to Christiansen et al. (2005), Benz and Trück (2006) distinguish two main drivers that should be taken into consideration when modeling EUA price dynamics, namely policy and regulatory issues as well as market fundamentals directly affecting supply and demand. The authors suggest applying a GARCH and Markov switching model in order to capture different phases of volatility in the data. Alberola et al. (2008) further characterize market fundamentals as well as identify structural breaks in April 2006 and October 2006. The first structural break occurs in April 2006 after disclosure of the 2005 verified emissions informing market participants about the net short/long positions of installations causing the EUA price to collapse, implying the framework is not stringent enough to incentivize CO_2 emissions reduction. The second structural break occurs after the European Commission's announcement regarding more stringent rules in the second trading phase form 2008 to 2012. In an analysis of the early trading period from 2015 to 2017, Daskalakis et al. (2009) find that the EUA price follows a non-stationary process, and discontinuous jumps are a prominent characteristic of the data consistent with Alberola et al. (2008) findings of structural breaks. Daskalakis et al. (2009) conclude that the price is most appropriately described by a Brownian motion augmented by a jump-diffusion process as proposed by Merton (1976). In addition, Daskalakis et al. (2009) highlight pricing differences for intra-phase and inter-phase derivatives due to the regulatory prohibition of banking between phases of the EU ETS. In a previous analysis Daskalakis and Markellos (2008) show both spot and futures EUA markets are not weak-form market efficient, highlighting the inadequacy of the initial trading period's regulatory framework.

The analysis by Borak et al. (2006) based on the initial trading period underlines the market expectations of price risk and high uncertainty of regulatory allocation plans reflected by significant convenience yields in future contracts beyond the initial phase of the EU ETS. However, Uhrig-Homburg and Wagner (2009) note that the link between spot and futures prices via the convenience yield approach is not useful for contracts that are maturing outside of the current trading period due to the previously described strict regulatory prohibition of banking and borrowing of EUAs outside of the current trading period. Nevertheless, Uhrig-Homburg and Wagner (2009) conclude that the cost-of-carry approach links spot and futures within the first trading period, and that futures markets are incorporating information first and then transfer the information to the spot market. In further studies Milunovich and Joyeux (2010) as well as Chevallier et al. (2010) do not find evidence to support a cointegration relationship between spot and futures contracts. However, Rittler (2012) concludes that mixed findings are due to low-frequency data and that he indeed finds an unambiguous cointegration relationship between spot and futures prices based on high-frequency intra day data. Rittler (2012) additionally confirms that futures markets are leading the price discovery. In a further application of high-frequency data, Conrad et al. (2012) show that the fractionally integrated asymmetric power GARCH specification as suitable in describes futures price dynamics, implying the second conditional moment of the time series is characterized by long memory, power effects, and asymmetry.

As suggested in their previous research, Benz and Trück (2009) implement a Markov switching model and an AR-GARCH model to capture price dynamics of emission allowances. The authors find models with constant variance clearly show a poor fit compared to models with conditional variance, including the Markov switching model, which shows the best in-sample fit. Similar to the in-sample results, the conditional variance models also significantly outperform out-of-sample in terms of density forecasting. Benz and Trück (2009) conclude that the outperformance of conditional variance models is due to the emission allowance price's relationship with regulatory factors and fundamental variables. Although Daskalakis et al. (2009) implemented a different model, as previously mentioned, to capture the distributional characteristics of the log returns they too confirm the non-normality and heavy tails of the log return series. Benz and Trück (2009) argue adequate pricing and forecasting models need to incorporate issues such as shifts in price, non-normality or short periods of extreme volatility. Paolella and Taschini (2008) develop an innovative GARCH-type structure to approximate the conditional dynamics while applying a Pareto distribution to unconditional tails in order to appropriately capture the dynamics of the emission allowance spot market returns. In later studies Benschopa and López Cabrera (2014) apply a Markov regime switching model to the data of the second trading period and find results confirming findings from Benz and Trück (2009) highlighting volatility clustering, breaks in the volatility process and heavy-tailed distributions are well captured by the proposed model. Segnon et al. (2017)

expand the analysis to a Markov switching multifractal (MSM) model and data observed across the second and third trading period and find the MSM model outperforms simple GARCH, FIGARCH, and Markov switching GARCH models for most performance indicators and forecasting horizons.

Besides the covered literature, there are additional streams of research, including research regarding the interaction of the EUA market with other markets, particularly the energy market, and research pertaining to the impact of emissions pricing on firm performance. However, the literature review will not expand to the aforementioned areas of research as the focus topic area of this thesis is in relation to the univariate stochastic properties of EUA prices.

3 Methodology

This section provides an overview of methodologies used in this thesis in order to analyze the time series properties of emission allowances, choose and estimate an appropriate mean and variance model, assess the model fit, and assess the model's forecasting accuracy. Both models for the mean (ARMA models) and the variance (GARCH model) are described in this section, noting the focus of the thesis is on the variance model due to the higher relevance in the context of the analyzed emission allowances time series. Before delving into the model methodology, the following section will provide an overview of relevant concepts and statistical test outlining the foundation for any further analysis. After that, the ARMA and GARCH models are outlined, and in the last subsection of this section, the forecasting accuracy measurement methods are described.

3.1 Stationarity, linear dependence, and normality

The concept of stationarity is important in time series analysis and is a prerequisite if one wants to draw an accurate conclusion from the models discussed in the following subsections. As Brooks (2002) outlines, the non-stationarity of a time series can strongly influence its behavior and properties, can lead to spurious regressions, and a regression's standard assumptions for asymptotic analysis are not valid. A time series is a strictly stationary process if the distribution of its values remains the same across time. In contrast, a weakly stationary process is one where the process has a constant mean, constant variance, and constant autocovariance structure. If a time series is found to be non-stationary, it can be differenced d times to become a stationary time series. In the context of financial time series, it is often the case that the price time series is considered non-stationary and that the first difference, the return series, is often considered stationary. This common transformation from prices to log returns is shown in Equation 1.

$$ln\left(\frac{S_t}{S_{t-1}}\right) = r_{log} \tag{1}$$

Visual inspection of the autocorrelation function (ACF) can be indicative of a unit root and hence non-stationarity. The autocorrelation is defined as

$$\rho_s = \frac{\gamma_s}{\gamma_0}\,,\; s = 0, 1, 2, .. \tag{2}$$

where γ_s is the autocovariance at lag s. While the ACF can suggest the existence of a unit root, a formal test is required for more certainty for the existence of such a unit root. One such test was developed by Dickey and Fuller (1979). The test's null hypothesis is that the series as described by Equation 3 contains a unit root, i.e., that $\phi = 1$ in

$$y_t = \phi y_{t-1} + u_t \tag{3}$$

while the alternative hypothesis is that $\phi < 1$ and hence stationary. For practical applications the test is set up for $\Delta y_t = y_t - y_{t-1}$ such that the regression becomes

$$\Delta y_t = \psi y_{t-1} + u_t \tag{4}$$

and hence the null is $\psi = 0$. Dickey and Fuller (1979) also developed extensions to this test that allow for a drift and deterministic time trend. A common criticism of the test, as outlined by Brooks (2002), is that its power is low, and it is difficult to correctly test processes with true ϕ close to 1.

For the time series models applied in this thesis, it is vital to determine the correct model specifications. Similarly to the non-stationarity detection, the autocorrelation can be assessed visually via the ACF. However, it is also useful to utilize a formal test in determining whether the autocorrelations at one or more lags are jointly statistically significantly different from zero. Ljung and Box (1978) developed such a formal test. The Ljung-Box statistic with the formal null hypothesis of

$$H_0:\; \rho_1 = \rho_2 = ... = \rho_{10} = 0$$

versus the alternative hypothesis

$$H_a:\; \rho_s \neq 0,\; for\; s \in \{1, 2, ..., 10\}\,,$$

and the Ljung-Box statistic is derived as follows:

$$Q_s^* = n(n+2)\sum_{k=1}^{s}\frac{\hat{p}_k^2}{n-k} \tag{5}$$

where n is the sample size, $\hat{p}_k^2$ is the sample autocorrelation function at lag k , and s is the number of lags. The Ljung-Box statistic Q_s asymptotically follows a χ^2 distribution with s degrees of freedom. The test is an extension and a more general form of the Box-Pierce test developed and named by Box and Pierce (1970).

In addition to testing for linear dependence, it is crucial for a variety of applications in time series analysis to test the normality. Normality is often an underlying assumption due to the desirable mathematical properties of the normal distribution and its high practical relevance. Hence it is also vital to test whether the assumption holds or not to, avoiding miss-specifications and incorrect decisions from statistical tests. Bera and Jarque (1981) developed the Jarque-Bera test defined as

$$JB = \frac{n}{6}\left(S^2 + (K-3)^2/4\right) \tag{6}$$

where n is the sample size, S is the sample skewness, and K is the sample kurtosis. The test using the properties of the normal distribution in regards to skewness and kurtosis to test whether the data is derived from a normal distribution. The test statistic is χ_2^2 distributed, and the null hypothesis is a joint hypothesis of

$$H_0:\ S = (K-3) = 0$$

as the expected sample skewness and excess kurtosis of the normal distribution are zero.

The following subsections will provide an overview of relevant mean and variance models in which the above concepts are embedded.

3.2 ARMA models

The ARMA models are widely used in time series analysis, combining autoregressive and moving average processes. The autoregressive model describes a model where the current value y_t depends on previous values of the time series and an error term. The moving average model is a linear combination of white noise error terms, meaning the current value y_t depends on the previous and current values of the error terms. The combination of the two types of models are ARMA models, defined as

$$y_t = \mu + \sum_{i=1}^{q} \phi_i u_{t-i} + \sum_{j=1}^{p} \theta_j y_{t-j} + u_t \tag{7}$$

with $E(u_t) = 0$, $E(u_t^2) = \sigma^2$, $E(u_t, u_s) = 0$, and $t \neq s$. The determination of ARMA(p,q) orders can be achieved via a graphical analysis of the ACF and partial autocorrelation function (PACF). The PACF is the correlation between lag s and the current lag while controlling for the correlation of the lags in between lag s and the current lag. Visual inspection of the ACF and PACF can be accompanied by information criteria that quantify the ranking of different model specifications. Popular information criteria are Akaike's information criteria (AIC) as defined by Akaike (1974) and the Bayesian information criteria (BIC) as defined by Schwarz (1978). The two information criteria are respectively defined as

$$AIC = ln(\hat{\sigma}^2) + \frac{2k}{n} \tag{8}$$

$$BIC = ln(\hat{\sigma}^2) + \frac{k}{n} ln(n) \tag{9}$$

where k is the number of parameters, n is the sample size, and $\hat{\sigma}^2$ is the residual variance. The information criteria will not necessarily give the same results as they impose a varying penalty for additional parameters in the model specification. However, the combination of ACF and PACF, as well as the usage of the information criteria, provide a broad picture to make the most appropriate model choice when facing noisy real data.

3.3 GARCH models

While ARMA models provide for the mean to follow an autoregressive and moving average process, this class of time series models is linear in nature. As Brooks (2002) points out, many financial time series are characterized by features that warrant the motivation to consider non-linear models. Such features include leptokurtosis, volatility clustering, and leverage effects. Leptokurtosis describes the tendency of financial asset returns to exhibit excess kurtosis and fat tails. Volatility clustering refers to the effect that a time series exhibits autocorrelation in the second moment, implying volatility occurs in clusters such that there distinct periods of high volatility and periods of low volatility. Leverage effects in financial data explain the tendency for a higher increase in volatility after a negative return shock compared to a positive return shock of the same magnitude. The previously mentioned ARMA models are not modeling the variance and thus assume homoscedasticity. This feature deems ARMA models unsuitable when considering a time series that exhibits heteroskedasticity. Estimating a model under the assumption innovations are homoscedastic while the time series is in fact implying the innovations are heteroscedastic, could lead to wrong standard error estimates.

Engle (1982) developed the first model addressing the heteroskedasticity in time series with his ARCH(p) model, where ARCH stands for “autoregressive conditionally heteroskedastic”. The ARCH model is defined as

$$y_t = \mu + u_t \tag{10}$$

$$u_t = \varepsilon_t \sigma_t \quad \varepsilon_t \sim N(0,1) \tag{11}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 \tag{12}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, ε_t are i.i.d. with zero mean and unit variance. The ARCH model captures leptokurtosis and volatility clustering due to the conditional volatility’s dependence on u_{t-i}. However, the model does have limitations in practical applications as outlined by Pagan (1996), specifically noting the inability of a parsimonious ARCH model to capture both height and shape dimensions of the ACF and hence fully describe the time series’s features. In addition, the non-negativity assumption for α_i is not practical for most financial time series, particularly if i is large, the probability of one or more negative parameter estimates is high, and the conditional variance could be negative, which is a meaningless result.

The Generalised ARCH (GARCH) model is overcoming limitations of the ARCH model and is widely used for financial time series applications. Both Bollerslev (1986) and Taylor, S (1986) independently developed the GARCH model. Assuming y_t and u_t as defined in Formula 10 and 11 respectively, the conditional variance under the GARCH (p,q) model is defined as:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{13}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, and $\beta_j \geq 0$ and $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$. The conditional variance process is stationary, and the conditional variance process is positive and consequently well defined if the aforementioned conditions are met. Considering a GARCH(1,1) model the unconditional variance is then defined as $Var(u_t) = \frac{\alpha_0}{1-(\alpha_1+\beta)}$. In contrast, if $\alpha_1 + \beta \geq 1$ the variance process is non-stationary, the unconditional variance is not defined, and such a model specification has undesirable characteristics not suitable for practical application. The model is dependent upon prior residuals and the prior conditional variance, making it synonymous with an ARMA(1,1) model of the squared residuals u_t^2. Compared to ARCH models, GARCH models are more suitable for most financial time series data as a more parsimonious model can be applied to capture the desired data properties.

Although the GARCH model is widely used, it does not account for leverage effects often observed in financial time series. Both the EGARCH as defined by Nelson (1991) and the GJR-GARCH as defined by Glosten et al. (1993) are known as asymmetric GARCH models, capturing similar asymmetry features.

GJR-GARCH is part of the asymmetric power GARCH (APGARCH) class of models that are characterized by including a power coefficient γ and leverage coefficient δ . The GJR-GARCH model has a power coefficient $\gamma = 2$, which reduces the model to the standard GARCH case of dependence upon lagged squared innovations and lagged conditional variance but including a leverage coefficient. Assuming y_t and u_t as defined in Formula 10 and 11 respectively, the conditional variance under the GJR-GARCH (p,q) model is defined as

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{q} \delta_i I[u_{t-i} > 0] u_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{14}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, and $\beta_j \geq 0$ and $\sum_{i=1}^{q} \alpha_i + \frac{1}{2} \sum_{i=1}^{q} \delta_i + \sum_{j=1}^{p} \beta_j < 1$. Additionally to the standard GARCH model, the indicator function is defined as $I = 1\, if\; u_{t-i} < 0 \quad and \quad I = 0\, if\; u_{t-i} < 0$ and the leverage coefficient $-1 < \delta < 0$. The conditional variance process is stationary, and the conditional variance process is positive and consequently well defined if the aforementioned conditions are met. The lagged innovation u_t is the decision variable in the indicator function, implying that the additive leverage term will only apply in cases of a negative u_t as the term is otherwise zero given multiplied by a zero indicator function. Consequently, the GJR-GARCH is useful under the hypothesis that negative shocks induce higher volatility than for positive ones with equal magnitude.

In comparison to the APGARCH family of GARCH models and the simple GARCH model, the EGARCH model is modeling the log of the conditional variance rather than the conditional variance. That distinction naturally implies that the non-negativity constraint is relaxed. The EGARCH model is featuring a leverage effect that can capture both asymmetric positive or negative shocks. The EGARCH model is defined by

$$y_t = \mu + u_t \tag{15}$$

$$u_t = z_t \sigma_t \tag{16}$$

$$log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \left[\frac{|u_{t-i}|}{\sqrt{\sigma_{t-i}^2}} - E\left\{ \frac{|u_{t-i}|}{\sqrt{\sigma_{t-i}^2}} \right\} \right] + \sum_{i=1}^{q} \delta_i \left(\frac{u_{t-i}}{\sqrt{\sigma_{t-i}^2}} \right) + \sum_{j=1}^{p} \beta_j log(\sigma_{t-j}^2) \tag{17}$$

where the expected value $E\left\{\frac{|u_{t-i}|}{\sqrt{\sigma_{t-i}^2}}\right\}$ depends on the distribution of z_t. For the standard case of a normal distribution, the expected value is defined as

$$E\left\{ \frac{|u_{t-i}|}{\sqrt{\sigma_{t-i}^2}} \right\} = E\left\{|z_{t-i}|\right\} = \sqrt{\frac{2}{\pi}}$$

and for the case of a student's t-distribution with ν degrees of freedom the expected value is defined as the following.

$$E\left\{ \frac{|u_{t-i}|}{\sqrt{\sigma_{t-i}^2}} \right\} = E\left\{|z_{t-i}|\right\} = \sqrt{\frac{\nu - 2}{\pi}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$$

The specification of an alternative distribution for the innovation distribution is not unique for the EGARCH model but can be applied to the aforementioned ARCH and GARCH models. However, only for the EGARCH does the choice of innovation distribution directly flow into the estimation function. While the normal distribution is the standard choice across model specification, the use of other distributional assumptions may be more appropriate for a specific set of data. The student's t-distribution is a more appropriate choice if the GARCH process is not fully capturing the leptokurtic properties of the data. Hence, if the innovation series of a GARCH process exhibit excess kurtosis and fat tails, a model specification with t-distributed innovations may be more appropriate.

3.4 Forecasting accuracy measures

The out-of-sample performance of the calibrated models is assessed via point forecast as well as density forecast accuracy measures. The forecasted returns are compared to the observed returns to determine the forecast errors and subsequently the accuracy measures. Common point accuracy measures, the mean squared error (MSE), and the mean absolute error (MAE), are employed to determine which model performs best. The MSE and MAE are defined as

$$MSE = \frac{1}{T-(T_1-1)} \sum_{t=T_1}^{T} (y_t - y_{f,t})^2 \tag{18}$$

$$MAE = \frac{1}{T-(T_1-1)} \sum_{t=T_1}^{T} |y_t - y_{f,t}| \tag{19}$$

where T is the total sample size, $T_1$is the first observation of the out-of-sample period, y_t is the realized return in time t and $y_{f,t}$ is the forecasted return in time t.

While point forecast accuracy measures are useful in determining out-of-sample performance, they do not paint the full picture. In contrast, density forecast accuracy measures provide a better understanding of the full density of the out-of-sample performance, particularly relevant for practical risk management applications. There are different approaches to achieve this goal as discussed by Christoffersen (1998), Bollerslev (1986), Diebold et al. (1998), and Crnkovic and Drachman (1995). This thesis replicates the application Benz and Trück (2009) used as suggested byDiebold et al. (1998) and Crnkovic and Drachman (1995). The accuracy of the density forecasts is assessed by utilizing distributional tests measuring the difference between an empirical and theoretical distribution. The test set up requires i.i.d variables with a comparable distribution and hence the forecasted variables need to be transformed. As Rosenblatt (1952) shows, if the loss distribution is correctly specified, the transformation

$$x_t = \int_{-\infty}^{y_t} \hat{f}(u)du = \hat{F}(y_{l,t}) \tag{20}$$

is resulting in i.i.d. and uniformly distributed x_t. In Equation 20 $y_{l,t}$ are the ex-post loss forecasts and $\hat{f}(\bullet)$ is the forecasted loss density. Importantly, the transformation holds independent of the underlying distribution of $y_{l,t}$. Once the x_t are determined, tests such as Kolmogorov-Smirnov (KS) and Kuiper test can be employed in order to check the data for uniformity. Both the Kuiper and KS tests are testing the null hypothesis that an empirical cumulative distribution function is equal to a theoretical cumulative distribution function. The KS test statistic is defined as

$$KS = max[|\hat{F}(x) - G(x)|]$$

and the Kuiper test statistic is defined as

$$\begin{aligned} D_{Kupier} &= D^+ + D^- \\ D^+ &= max[\hat{F}(x) - G(x)] \\ D^- &= max[G(x) - \hat{F}(x)] \end{aligned}$$

where $\hat{F}(x)$ is the empirical cumulative distribution function and $G(x)$, and is the theoretical distribution function. In this context $\hat{F}(x)$ is representing the probability integral transforms of the one day ahead forecasts and $G(x)$ represents the uniform cumulative distribution function. In addition to the statistical test, Diebold et al. (1998) suggest assessing the histograms of the probability integral transforms. The histograms should describe a uniform distribution, and the shape of the histograms can provide additional feedback on the models' shortcomings.

4 Empirical results

4.1 Data

The data which this thesis is resting upon is obtained from the European Energy Exchange (EEX) through Bloomberg. The EEX which is one of the major spot exchanges in which EUAs can be traded and is traded with significant enough liquidity to provide a non-distorted liquid spot price. Each spot contract for EUAs permits the holder to emit one ton of carbon dioxide or one ton of a carbon dioxide equivalent, within the meaning of the EU ETS. The secondary market tick size is 1000 EUAs. The data comprises daily EUA spot prices, and the considered time period is part of the third trading period and spans from January 01, 2016 to December 31, 2019. The data period was chosen to reflect the latest developments and regulatory changes and to avoid noisy log returns in a low price level environment in the early years of the third trading period while ensuring a sufficient sample size. The data from January 01, 2016 to December 31, 2017 is used to calibrate the models, i.e. the in-sample data set, while the remaining period from January 01, 2017 to December 31, 2019 is used to assess out-of-sample performance. Figure 6 shows the daily spot prices of the third trading period. From visual inspection, it is clear that the market experienced low variability in the level of prices within the majority of the third trading period, with prices having remained in the single-digit range until 2018. For the purpose of using the data for the proposed time series models, the data needs to be stationary. Visual inspection of the price graph as well as the ACF and PACF in Figures 4 and 7, respectively, lead to the conclusion that the price time series is not stationary. The persistent ACF shows signs of a unit coefficient on the lagged dependent variable, implying that shocks do not die away. The PACF shows only the first lag is significant, implying no moving average term is appropriate and in connection with the ACF show the classic characteristics of a non-stationary time series. The augmented Dickey Fuller test with results in Table 2 confirms the conclusion from the visual evidence.

Consequently, the time series is transformed by deriving the log return series according to Formula 1. Figure 8 shows he transformed data series in the full third trading period while Figure 10 Panel (a) shows the log returns for the considered time period for calibration and out-of-sample testing. In addition, Table 1 shows summary statistics for the full considered period, the in-sample period, and out-of-sample period. The stationarity is confirmed by the augmented Dickey Fuller test with results in Table 2. The log returns series in the different considered periods show the mean is very close to zero not larger than 25bps in all periods and not significantly different from zero given the null hypotheses under the student's t-test can not be rejected at the 5% significance level. The minimum of the log return series could be observed in September 2018 with a negative 19.5% return, and the maximum return of 14.5% occurred in December 2016. In the calibration period, the log returns are right-skewed, exhibiting skewness of 0.16, while the out-of-sample series is left-skewed with -0.78 skewness. However, only the out-of-sample skewness is significantly different from zero at the 5% significance level. Both the in-sample and out-of-sample periods are characterized by highly significant excess kurtosis. Figure 9 shows the histograms of log returns for the three periods and the fitted normal distribution and t-distribution. In line with the descriptive statistics, one can see the empirical log return distribution is peaked around the mean and has fat tails. The fitted t-distribution appears to be better in describing the empirical distribution compared to the normal distribution. The Jarque-Bera test for normality confirms the non-normality of the data in the three considered time periods. Figure 10 show the log returns, squared log returns, and absolute log returns showing returns occurring in bursts indicating heteroskedasticity. The leptokurtosis, volatility clustering, and leverage effects in the data series indicate a simple normal distribution is not sufficient to model data, and hence alternative models shall provide a better fit.

4.2 Time series models

The analysis of the log returns suggests that GARCH models may be appropriate to fit the data. Therefore, in-sample fit and out-of-sample performance are assessed for a simple GARCH model, a GJR-GARCH, and an EGARCH model. Based on the results in the data analysis, it is clear that the time series is non-normal and shows strong signs of leptokurtosis. Therefore, the error terms in the GARCH model are assumed to follow a t-distribution instead of a Gaussian distribution. In addition, results from the GARCH models are benchmarked on a t-distribution of the log returns. The benchmark model is not explicitly modeling the volatility of log returns but only the mean equation. Both mean and volatility equations need to be assessed for adequacy. In the first step, the appropriateness of an ARMA model for the mean equation is assessed. The ACF and PACF of log returns in Figure 11 indicate there is little evidence in the data to support an AR or MA or ARMA model for the mean process. However, given the third lag appears to be slightly significant, the application of an analytical test is appropriate to confirm the intuitive graphical evidence. The Ljung-Box test statistic for linear dependence, as defined in Equation 5, is statistically not significant for the joint test for ten lags, as shown in Table 3. Therefore, the null hypothesis of zero autocorrelation at all lags cannot be rejected. In an additional test, the information criteria for various combinations of model specification, p and q , are determined, and results are illustrated in Table 5. While the AIC suggests a model specification of ARMA(2,2), the BIC suggests ARMA(0,0). Overall, the results are mixed. However, in combination, the ARMA(0,0) is most appropriate supported by the lacking evidence for significant autocorrelation as determined by the LB test, the graphical indication of the ACF and PACF as well as the BIC. In addition to the combined picture of the test and figures, a parsimonious model is desirable. On that basis the mean equation is describes as

$$y_t = \mu + u_t \tag{21}$$

$$u_t = \varepsilon_t \sigma_t \tag{22}$$

where $\varepsilon_t \sim t(0, \sigma_t^2, df)$. The mean equation serves as the basis for all considered models in this thesis. While all GARCH models specifically add a variance equation, the naive model does not. Hence the naive model assumes $\sigma_t = \sigma$, i.e. the model implies an unconditional variance.

4.3 In-sample results

In the following, the in-sample results for the four considered models are discussed. A naive model is chosen to benchmark the results of the GARCH models. The naive model is fitting a t-distribution on the log return series resulting in parameter estimates of $\mu = -0.00042$ and $\sigma^2 = 0.0011$ and degrees of freedom of $df = 5.19$. Hence the model is fully described by $y_t = \mu + u_t$ where $u_t \sim t(0, \sigma^2, df)$. The parameter estimates are derived via maximum likelihood estimation for all models in this thesis. The residual series of the naive model appears to exhibit non-constant variance illustrated by the ACF and PACF of the squared residual series in Figure 12. In addition, the Ljung-Box test results for the joint test for ten lags, as shown in Table 3, is rejecting the null of zero autocorrelation in the squared residual series with a p-value close to zero. The ARCH test, as defined by Engle (1982), is confirming the presence of heteroskedasticity in the time series. The ARCH test results for the standardized residual series are shown in Table 3. Based on the results, a GARCH model is an appropriate choice to improve the in-sample fit compared to the naive model. In line with the prevalent usage in the academic literature, a parsimonious GARCH(1,1) model is fitted. The variance equation is therefore described by

$$u_t = \varepsilon_t \sigma_t \quad \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{23}$$

where ε_t is i.i.d. and $\varepsilon_t \sim t(0, \sigma^2, df)$. The parameter estimation results are presented in Table 6. The GARCH coefficient is highly significant with a p-value close to zero, and the ARCH coefficient is significant at the 5% significance level. Both constant terms are not significant, while the degrees of freedom are highly significant. Similarly to other financial time series, the sum of α_1 and β_1 is close to 1; in this case approximately 0.95, implying a highly persistent volatility structure. In the context of financial time series, the ARCH coefficient α_1 is high with an estimate above 0.1 while the GARCH coefficient is at the low end with 0.85. Both features are indicating a spiky volatility structure observable in time series of an active market environment. Figure 13 displays the innovations and conditional standard deviation of the GARCH(1,1) model. The conditional standard deviation clearly shows significantly varying levels of volatility occurring in spikes. Noting the volatility decreased towards the end of 2017 with only a small spike in October. Overall the volatility in the in-sample period spans from quiet periods with low volatility in the 0.02 to 0.025 range and more active periods in which the volatility increases to up to 0.06. While the GARCH model is capturing the volatility clustering and leptokurtosis of the time series, it does not capture any leverage effects. In order to verify whether asymmetric leverage effects adding additional explanatory power to the model a GJR-GARCH(1,1) model is calibrated and the variance equation is defined as

$$u_t = \varepsilon_t \sigma_t \quad \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \delta_1 I[u_{t-1} > 0] u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{24}$$

where ε_t is i.i.d. and $\varepsilon_t \sim t(0, \sigma^2, df)$. Similarly to the GARCH model for the GJR-GARCH model, the ARCH and GARCH lag specification at 1 follows standard practice in the academic literature. The estimation results are presented in Table 7. The derived parameter estimates for the GJR model are similar to the GARCH parameter estimates in terms of magnitude. The significance of the estimates follows a similar pattern with the GARCH coefficient being highly significant, while the ARCH coefficient is only significant at the 10% significance level, and the mean and variance constants are not significant. The additional leverage coefficient δ in the GJR model is not statistically significant, hinting at the possibility that there is an inability of the additional parameter to improve the model. Compared to the GARCH model, the GJR-GARCH volatility structure is spikier and has maximum volatility of 0.062 while the minimum is in line with the GARCH model's minimum.

Finally a EGARCH(1,1) model is calibrated to the data. The EGARCH model differs in the parameter estimation as the model is not modeling σ^2 but rather $log(\sigma^2)$. The EGARCH(1,1) model is defined as

$$u_t = z_t \sigma_t \quad log(\sigma_t^2) = \alpha_0 + \alpha_1 \left[\frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - E\left\{ \frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} \right\} \right] + \delta_1 \left(\frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} \right) + \beta_1 log(\sigma_{t-1}^2) \tag{25}$$

with

$$E\left\{ \frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} \right\} = E\left\{ |z_{t-1}| \right\} = \sqrt{\frac{\nu - 2}{\pi}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$$

where z_t is i.i.d and $z_t \sim t(0, \sigma^2, \nu)$. Table 8 displays the parameter estimates. The ARCH and GARCH coefficients are both highly significant, while the other coefficients, including the leverage parameter, are not statistically significant. The leverage coefficient is negative, implying negative shocks have a more substantial impact on volatility than positive ones of the same magnitude. This finding is in line with the assumptions under the GJR-GARCH model. The volatility structure of the EGRACH model is illustrated in panel c of Figure 13. The volatility structure is less spiky and has a maximum value of 0.57, the lowest out of the three variance models.

Results of the Ljung-Box test for serial autocorrelation of the respective squared residual series are presented in Table 10. All residuals from GARCH models can not reject the null of zero autocorrelation implying the three GARCH models have appropriately captured autocorrelation in the data. This result also confirms the appropriateness of a small order in GARCH models for this time series as expected and in line with wide application for financial time series. In contrast and as previously shown, the squared residuals series of the ARMA(0,0) process produces squared residuals with significant autocorrelation. Similarly, the ARCH test for heteroskedasticity, presented in Table 11, confirms the Ljung-Box test results implying GARCH models capture the heteroskedasticity while the naive models does not. The test results confirm by visual inspection of the ACF and PACF of the residuals series, illustrated in Figure 14. The log-likelihood alongside the information criteria AIC and BIC of the four considered models is presented in Table 9. All models incorporating a conditional variance specification clearly outperform the naive t-distribution model in terms of log-likelihood and information criteria. While the EGARCH model exhibits the highest log-likelihood, all three GARCH models are very close in terms of in-sample fit. The simple GARCH performs best according to both information criteria. The lower amount of parameters is making up for the slightly lower log-likelihood compared to the asymmetric models. Overall, none of the models incorporating a conditional variance process outperforms, but all models fit better to the EUA log return time series than the naive model. Therefore, one can conclude that in-sample volatility clustering and leptokurtosis is an essential feature in the data, while leverage appears to be less vital. In the following subsection, the out-of-sample performance is assessed.

4.4 Forecasting results

The out-of-sample performance of the four considered models is determined by comparing of one day ahead point forecast accuracy as well as assessing the density forecasts. The out-of-sample period spans from January 01, 2018 to December 31, 2019. The models are assessed based on return forecasts rather than volatility forecasts due to the inherent difficulty in specifying an appropriate benchmark for the forecasted volatilities. The forecasts are based on a fixed, recursive, and rolling window of data-samples considered for estimating the parameter coefficients. The fixed window implies only the in-sample data is used to estimate the coefficients, and there is no re-estimation. This approach can be interpreted as a benchmark as one would expect the re-estimation of parameters, incorporating more recent data points, would improve the forecasting accuracy. The recursive window uses all in-sample data points, and the amount of data points used is expanding with each re-estimation iteration of the one-step-ahead forecast. The rolling window is re-estimating at each iteration and statically moving the estimation window such that the amount of data points used is staying constant. In contrast, the window moves with each iteration. Therefore, both recursive and rolling windows are incorporating the most recent data to estimate parameters compared to the fixed window only using in-sample data.

The point forecasts accuracy is measured by the mean squared errors (MSE) and the mean absolute errors (MAE). Table 12 presents the mean errors for each considered model and re-estimation window approach. Overall, the results are very close; however, the simple GARCH model has the lowest errors across forecasting windows and accuracy measures. As expected, the fixed window shows the highest errors in all four considered models, implying that re-estimation is beneficial to model accuracy. The recursive window is performing best according to MSE, and the rolling window is outperforming in terms of MAE. The naive model exhibits the highest errors in all categories, while the difference in terms of magnitude is still very small. The small difference highlights the need for a more sophisticated approach to measuring the models' out-of-sample performance. Particularly in terms of practical relevance, it is vital to gain an understanding of not only the models' return forecasting accuracy but also their risk forecasting accuracy. In order to achieve a more in-depth understanding of the models' forecasting accuracy, the density forecasts are assessed. As previously outlined in subsection 3.4, there are several approaches to achieve this measurement. The fundamental transformation, according to Rosenblatt (1952), is the basis for the distributional test employed in this thesis. The histograms of the probability integral transforms of the four considered models are presented in Figure 15. A perfect model would display a flat line representing the uniform distribution implied by the transformation. While all models do show a tendency to higher peaks in the middle section, it appears the GARCH models are less peaked than the naive model. This tendency for concentrated frequencies in the middle quartiles implies more often than not that the forecasted confidence intervals are too wide. The fact that the naive model is exhibiting such behavior is not surprising as the variance is unconditional and hence does not consider volatility clustering. Visually it is difficult to see a clear best performing model among the GARCH models. Interestingly, all models, including the naive model, show a high frequency of the lowest distribution bucket on the left side of the graphs. At the same time, the highest distribution bucket on the right side of the graphs is infrequent across models. This concentration implies the models failing to capture fat tails on the left side of the return distribution, i.e., negative returns. These results align with the findings in the descriptive statistics, showing the out-of-sample returns are left-skewed with a high statistical significance. While one would expect the asymmetric models to show a less pronounced feature, they in fact show higher frequencies in the lowest distribution bucket. As discussed in the in-sample analysis, the estimated leverage coefficients are relatively small and not statistically significant in the in-sample period, which extended throughout the re-estimation in the out-of-sample period implying a lacking ability of the models to capture the data features appropriately. Figure 16depicts the 95% confidence bands and the observed returns underlying the findings mentioned above from the histograms in Figure 15. The GARCH model shows the most pronounced reaction to high volatility in returns, particularly in the third and fourth quarter of 2018. The flat confidence intervals of the naive model represent the lacking modeling of conditional variance.

While the visual inspection provides a meaningful interpretation of the results, it is difficult to assess the best model out-of-sample performance ultimately, and hence a quantification with the help of a statistical test is a useful tool. Table 13 provides a summary of the test results of the KS and Kuiper test for uniformity of the probability integral transforms of the one day ahead forecasts. High p-values imply the underlying distribution of the probability integral transforms is close to the uniform distribution. Hence a p-value of one would imply the histograms in Figure 15 were completely flat. An additional visual interpretation of the data is illustrated in Figure 17. The Figure represents the empirical cumulative distribution function of the probability integral transforms, and the theoretical uniform cumulative distribution function. The statistical tests specifically use the cumulative distribution functions and determine the goodness of fit via distance between the two functions. The results in Table 13 show that the naive model is performing worst by a large margin and is rejected at the 10% and 1% significance level at the KS and Kuiper test, respectively. This result can be easily inferred from panel d in Figure 17 as the gray line showing the empirical cumulative distribution function is clearly different from the blue line representing the theoretical uniform cumulative distribution function. The three GARCH models show better results both in the statistical test and also visually the empirical lines are closer to the theoretical ones. The asymmetric models markedly outperform the simple GARCH model, and the EGARCH model performs best overall for both the KS and Kuiper test. The simple GARCH model rejects the null of a uniform distribution at the 10% significance level for the Kuiper test while it does not reject the null under the KS test. On an aggregated basis, the EGARCH model appears to be performing best on an out-of-sample basis while noting the results are overall close. Notably, Berkowitz (2001) points out that both the Kuiper and KS tests have some instability in sample sizes below 1000 data points. Taking this into consideration, the difference between the GJR-GARCH model and the EGARCH model is indeed very small and likely not significant. Nevertheless, the difference between the GARCH models and the naive model is large enough to conclude there is a significant difference presumably not materially impacted by the sample size. Hence the models modeling the conditional variance are markedly outperforming the naive model in terms of out-of-sample density forecast accuracy.

4.5 Comparison with the existing literature

The existing literature shows varying results in different time periods. Looking at data from the pilot phase, Benz and Trück (2009) find that Markov-Switching models perform best, capturing different phases of volatility and heteroskedasticity in log returns. The authors suggest regime switching models and GARCH models, incorporating conditional variance equations, are appropriate to account for shifting uncertainty arising from regulatory, political, and fundamental drivers. Similarly, Daskalakis et al. (2009) highlight that prices in the pilot phase are characterized by jumps and are non-stationary. The authors also find skewness, leptokurtosis, and find non-normality for both prices and returns. The authors that are covering the pilot phase highlight the immaturity of the emissions market and the significant uncertainty driven by the regulatory framework. In a study analyzing data from the second trading period, Benschopa and López Cabrera (2014) apply a Markov switching GARCH model and confirming findings from the pilot phase, as shown by Benz and Trück (2009), highlighting volatility clustering, breaks in the volatility process, and a heavy-tailed distribution of emissions returns. Segnon et al. (2017) show that the Markov switching multifractal model is outperforming simple GARCH and Markov switching models based on data across the second and third trading period. The results of this thesis support the overarching notion that volatility clustering and leptokurtosis are a prominent feature in the data across all trading periods, including the latest market data.

5 Conclusion

This thesis examines the short term log return behavior of EUAs and focuses on the volatility dynamics in the recent market environment. The performance of three different GARCH models is assessed and benchmarked against a naive model. The results are useful for market participants in complementing their trading strategies, furthering their risk management, and providing enhanced decision tools for hedgers. The daily log returns are characterized by skewness, excess kurtosis, fat tails, and varying levels of volatility. All three GARCH models provide a better in-sample fit than the naive model, a simple t-distribution. The EGARCH model shows the highest log-likelihood, but all three GARCH models show close in-sample goodness of fit. The out-of-sample forecasting accuracy assessment shows similar results. While the point forecasts show only small superiority of the conditional variance models, the density forecast assessment clearly shows the naive model is inferior. The simple GARCH model shows the best point forecast results while fails to hold up performance when considering the density forecasts. The EGARCH model shows the best results in terms of density forecasts, marginally outperforming GJR-GARCH. The simple GARCH model and the naive t-distribution model fail to show statistically convincing results in regards to the density forecast assessment.

The results strongly support the use of conditional variance models in modeling the daily log returns of EUAs. Hence volatility clustering is a vital feature in the data, highlighting that the impact from regulatory framework developments and fundamental drivers of EUAs are creating short term changes in the volatility structure. The good performance of the EGARCH and GJR-GARCH models both in terms of in-sample fit and out-of-sample performance hint that leverage may be an essential feature in the time series. However, the notion is curtailed by the small magnitude and statistical insignificance of the leverage coefficient as well as the lacking overarching outperformance compared to the simple GARCH model. Both the descriptive statistics of daily EUA log returns and the EUA price graph show a change in EUA behavior in the out-of-sample period compared to the in-sample period. The log returns exhibit strong skewness in the out-of-sample period, and the price reaches new significantly higher levels compared to the majority of the third trading period. Therefore, it is imperative to continually re-assess the short term return and volatility dynamics and aim to capture the latest features in the data. This notion of development is mirrored in the existing literature showing varying GARCH, and regime switching models outperform depending on the considered time period. However, both the existing literature and this thesis conclude the existence of volatility clustering and leptokurtosis is prominent in the respective time series and highlights the importance of volatility modeling for practical applications. Particularly in the context of risk management applications, the density forecast results indicate that additional research may be appropriate in formulating a suitable model that incorporates the latest features of the time series.

References

Aatola, P., M. Ollikainen, and A. Toppinen (2013). Price determination in the EU ETS market: Theory and econometric analysis with market fundamentals. *Energy Economics 36*, 380–395.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*(6), 716–723.

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

Benschopa, T. and B. López Cabrera (2014). Volatility modelling of CO_2 emission allowance spot prices with regime-switching GARCH models. Technical report, Humboldt University of Berlin.

Benz, E. and S. Trück (2006). CO_2 emission allowances trading in Europe – specifying a new class of assets. *Problems and Perspectives in Management 4*(3), 30–40.

Benz, E. and S. Trück (2009). Modeling the price dynamics of CO_2 emission allowances. *Energy Economics 31*, 4–15.

Bera, A. K. and C. M. Jarque (1981). An efficient large–Sample test for normality of observations and regression residuals. *Australian National University Working Papers in Econometrics 40*.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics 19*(4), 465–474.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Borak, S., W. K. Härdle, S. Trück, and R. Weron (2006). Convenience yields for CO_2 emission allowance futures contracts. *Working Paper, Humboldt University of Berlin*.

Box, G. E. P. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association 65*, 1509–1526.

Brooks, C. (2002). *Introductory econometrics for finance*. Cambridge University Press.

Chesney, M. and L. Taschini (2012). The endogenous price dynamics of emission allowances and an application to CO_2 option pricing. *Applied Mathematical Finance 19*(5), 447–475.

Chevallier, J. (2011). A model of carbon price interactions with macroeconomic and energy dynamics. *Energy Economics 33*(6), 1295–1312.

Chevallier, J. et al. (2010). A note on cointegrating and vector autoregressive relationships between CO_2 allowances spot and futures prices. *Economics Bulletin 30*(2), 1564–1584.

Christiansen, A. C., A. Arvanitakis, K. Tangen, and H. Hasselknippe (2005). Price determinants in the EU emissions trading scheme. *Climate Policy 5*(1), 15–30.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review 39*, 841–862.

Conrad, C., D. Rittler, and W. Rotfuß (2012). Modeling and explaining the dynamics of European Union allowance prices at high-frequency. *Energy Economics 34*, 316–326.

Crnkovic, C. and J. Drachman (1995). Quality control. *Risk 9*, 139–143.

Daskalakis, G. and R. N. Markellos (2008). Are the European carbon markets efficient. *Review of Futures Markets 17*(2), 103–128.

Daskalakis, G., D. Psychoyios, and R. N. Markellos (2009). Modeling CO_2 emission allowance prices and derivatives: Evidence from the European trading scheme. *Journal of Banking and Finance 33*, 1230–1241.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*, 427–431.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review 39*, 863–883.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*(4), 987–1007.

European Union (2003). Directive 2003/87/EC of the European Parliament and of the Council of 13 October 2003 establishing a system for greenhouse gas emission allowance trading within the Union and amending Council Directive 96/61/EC. *Official Journal of the European Union L275*, 32–46.

European Union (2009). Directive 2009/29/EC of the European Parliament and of the Council of 23 April 2009 amending Directive 2003/87/EC so as to improve and extend the greenhouse gas emission allowance trading scheme of the Community. *Official Journal of the European Union L140*, 63–87.

European Union (2014). Commission Regulation (EU) No 176/2014 of 25 February 2014 amending Regulation (EU) No 1031/2010 in particular to determine the volumes of greenhouse gas emission allowances to be auctioned in 2013-20. *Official Journal of the European Union L 56*, 11–13.

European Union (2015). Decision (EU) 2015/1814 of the European Parliament and of the Council of 6 October 2015 concerning the establishment and operation of a market stability reserve for the Union greenhouse gas emission trading scheme and amending Directive 2003/87/EC. *Official Journal of the European Union L 264*, 1–5.

European Union (2018). Directive (EU) 2018/410 of the European Parliament and of the Council of 14 March 2018 amending Directive 2003/87/EC to enhance cost-effective emission reductions and low-carbon investments, and Decision (EU) 2015/1814. *Official Journal of the European Union L 76*, 3–27.

Fehr, M. and J. Hinz (2006). A quantitative approach to carbon price risk modeling. *Institute of Operations Research, ETH Zurich*.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance 48*(5), 1779–1801.

Hammoudeh, S., D. K. Nguyen, and R. M. Sousa (2014). What explain the short-term dynamics of the prices of CO_2 emissions? *Energy Economics 46*, 122–135.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika 65*(2), 297–303.

Mansanet-Bataller, M., A. Pardo, and E. Valor (2007). CO_2 prices, energy and weather. *The Energy Journal 28*(3), 73–92.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics 3*, 125–144.

Milunovich, G. and R. Joyeux (2010). Testing market efficiency in the EU carbon futures market. *Applied Financial Economics 20*(10), 803–809.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*(2), 347–370.

Pagan, A. (1996). The econometrics of financial markets. *Journal of Empirical Finance 3*, 15–102.

Paolella, M. S. and L. Taschini (2008). An econometric analysis of emission trading allowances. *Journal of Banking and Finance 32*, 06–26.

Rittler, D. (2012). Price discovery and volatility spillovers in the European Union emissions trading scheme: A high-frequency analysis. *Journal of Banking and Finance 36*, 774–785.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics 23*, 470–472.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*, 461–464.

Segnon, M., T. Lux, and R. Gupta (2017). Modeling and forecasting the volatility of carbon dioxide emission allowance prices: A review and comparison of modern volatility models. *Renewable and Sustainable Energy Reviews 69*, 692–704.

Seifert, J., M. Uhrig-Homburg, and M. Wagner (2008). Dynamic behavior of CO_2 spot prices. *Journal of Environmental Economics and Management 56*(2), 180–194.

Taylor, S (1986). *Modeling financial time series*. John Wiley & Sons.

Uhrig-Homburg, M. and M. Wagner (2009). Futures price dynamics of CO_2 emission allowances: An empirical analysis of the trial period. *The Journal of Derivatives 17*(2), 73–88.

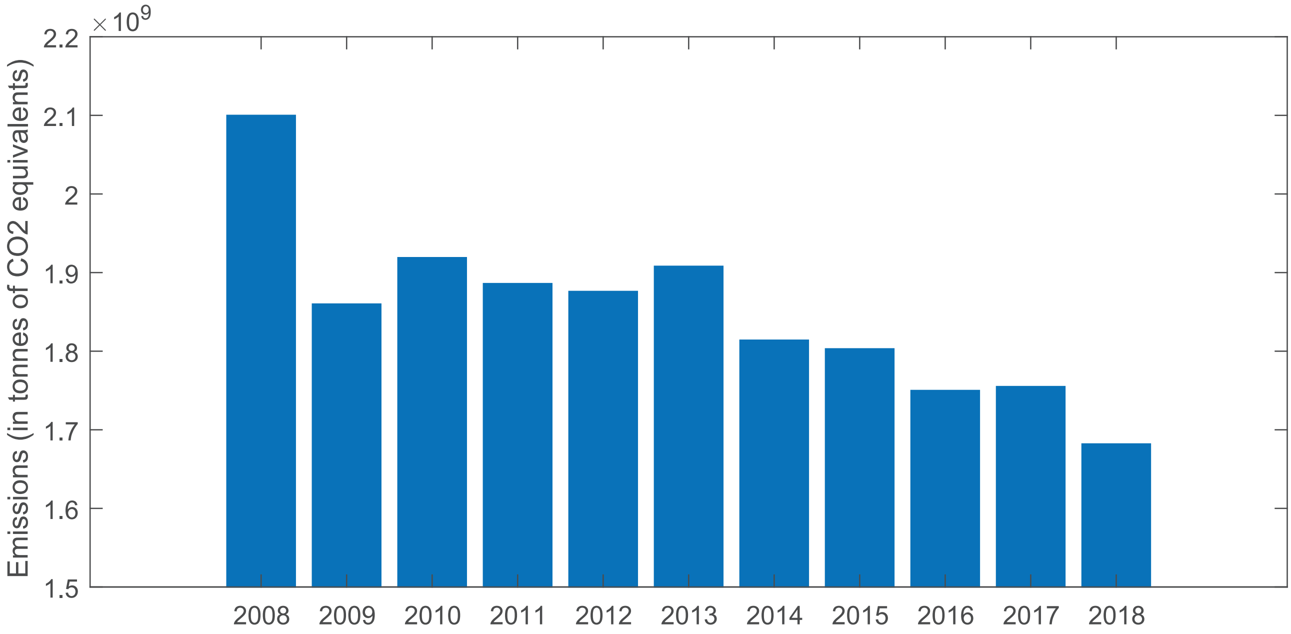


Figure 1: Annual verified emissions across the second and third phases of the EU ETS.

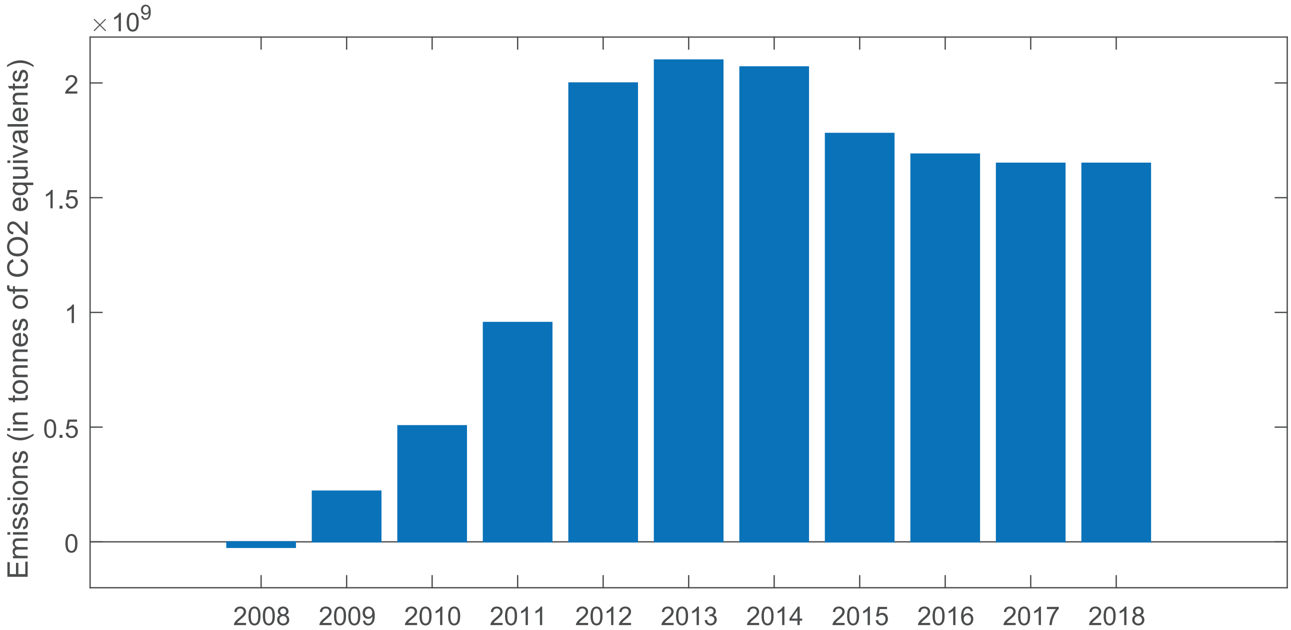


Figure 2: Development of the accumulated surplus across the second and third phases of the EU ETS.

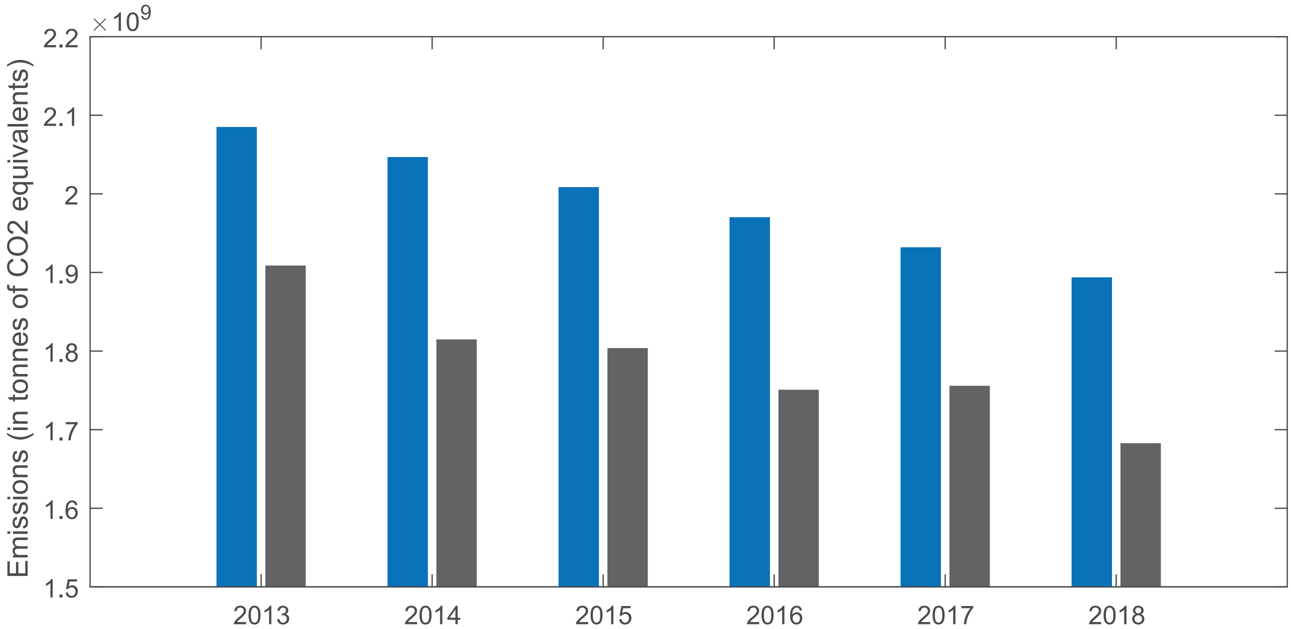


Figure 3: Annual emissions cap for installations and verified emissions in the third phase of the EU ETS.

Figure 4: Daily EUA spot prices from April 2013 to January 2020

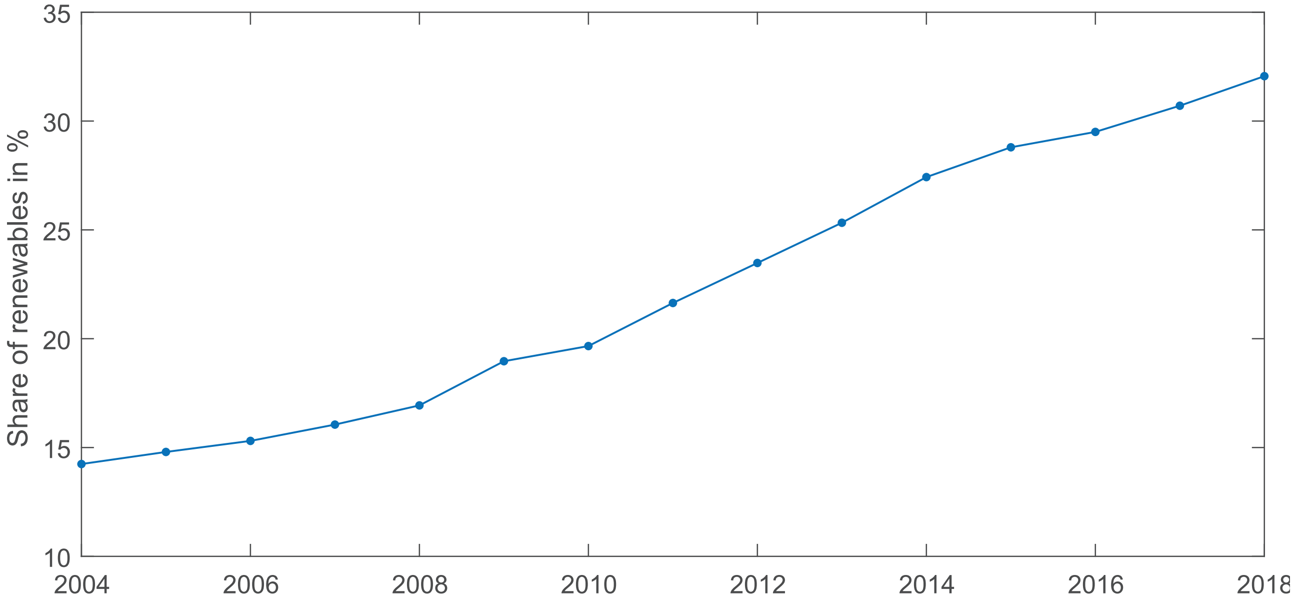


Figure 5: The share of renewable energy consumption in gross final energy consumption according to the Renewable Energy Directive in the EU (28 countries) from 2008 to 2018. The gross final energy consumption is the energy used by end-consumers plus grid losses and self-consumption of power plants. The presented data refers to renewable energy sources in electricity only.

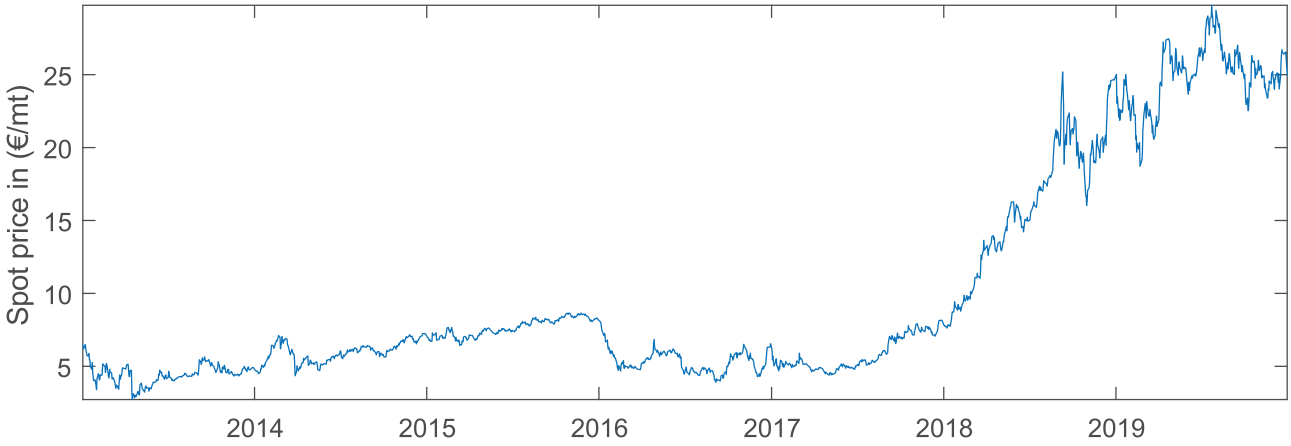


Figure 6: Daily EUA spot prices from January 01, 2013 to December 31, 2019

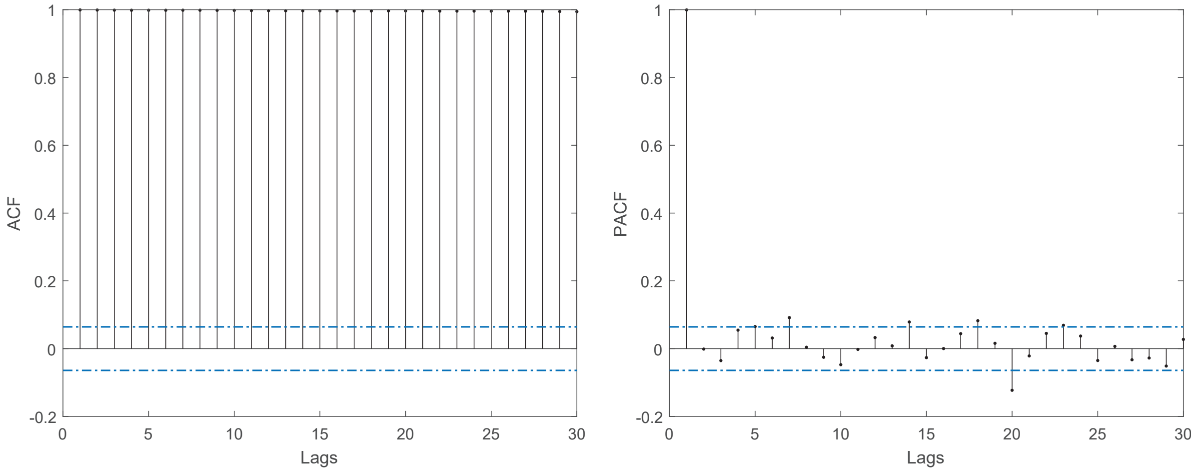


Figure 7: Sample autocorrelation function (left panel) and sample partial autocorrelation function (right panel) of the daily EUA prices with non-robust standard errors. The considered sample period spans from January 01, 2016 to December 31, 2019.

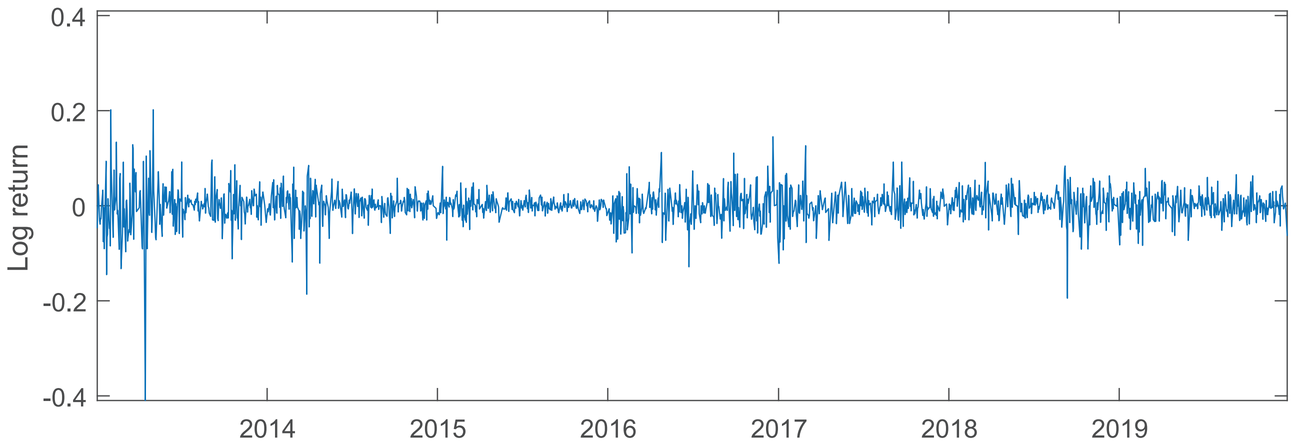
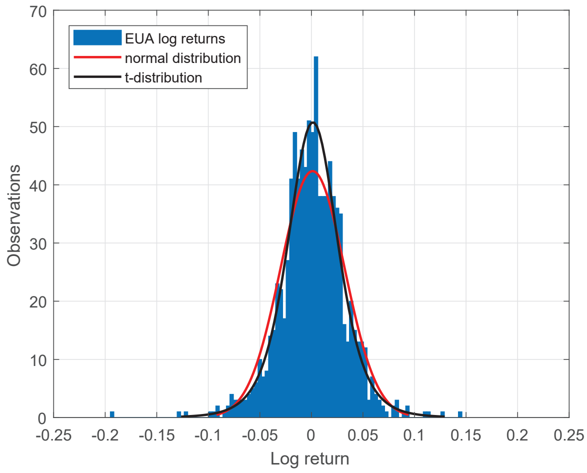
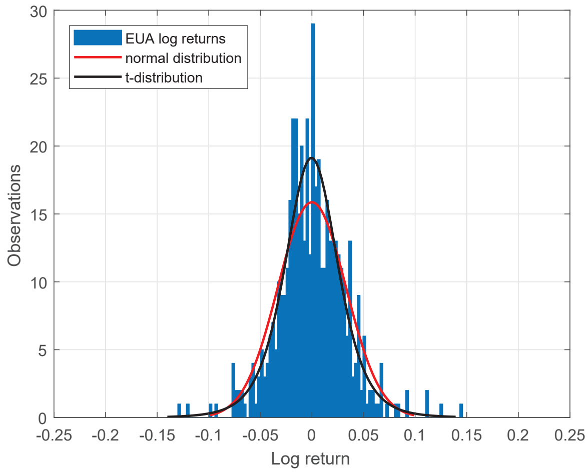


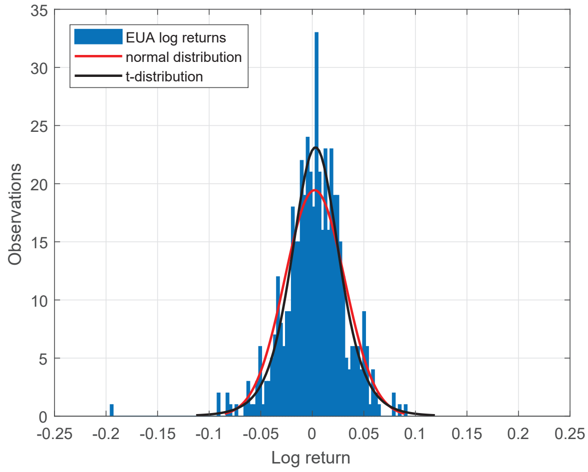
Figure 8: Daily EUA log returns from January 01, 2013 to December 31, 2019



(a) Full period

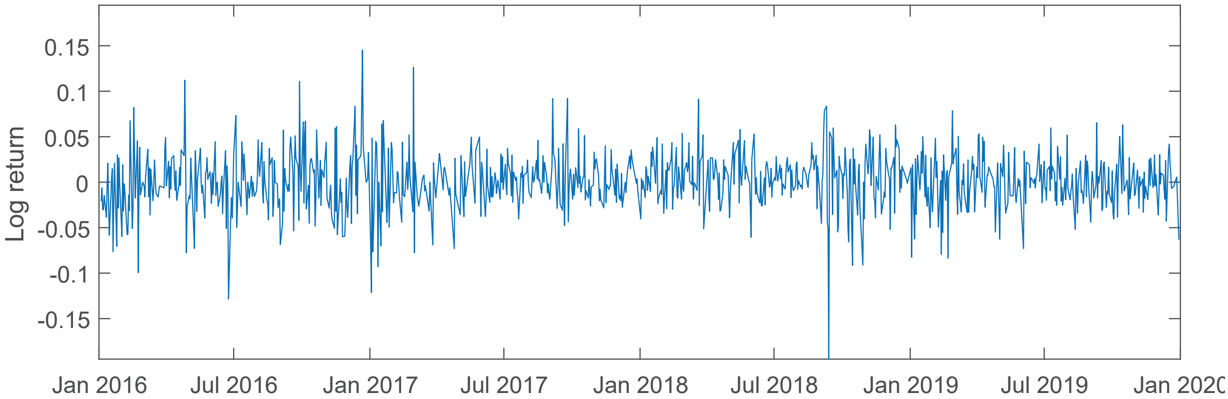


(b) In-sample period

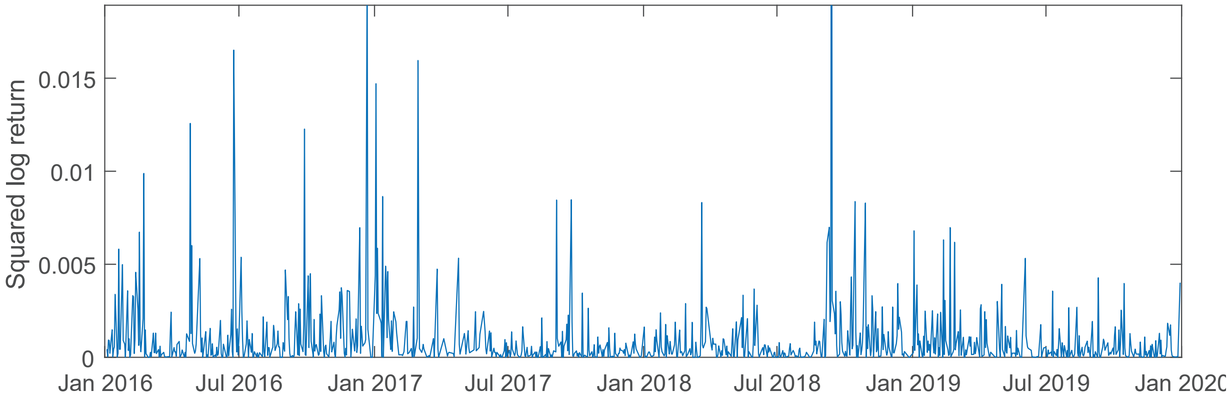


(c) Out-of-sample period

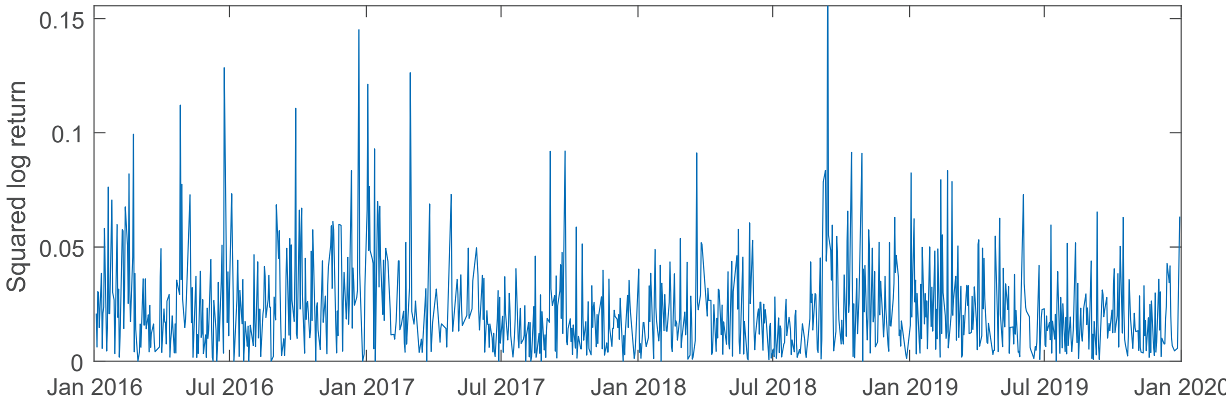
Figure 9: Histogram of daily EUA log returns for the full considered period from January 01, 2016 to December 31, 2019 (panel a), the in-sample period from January 01, 2016 to December 31, 2017 (panel b), and out-of-sample period from January 01, 2018 to December 31, 2019 (panel c) with fitted normal distribution and t-distribution.



(a) Log returns



(b) Squared log returns



(c) Absolute log returns

Figure 10: Daily EUA log returns (panel a), squared log returns (panel b), and absolute log returns (panel c) from January 01, 2016 to December 31, 2019

Table 1: Descriptive statistics and statistical tests for the daily EUA log return series for the full considered period from January 01, 2016 to December 31, 2019, the in-sample period from January 01, 2016 to December 31, 2017, and out-of-sample period from January 01, 2018 to December 31, 2019. Mean, skewness, and excess kurtosis are tested for zero mean. Jarque-Bera designates the empirical statistic of the Jarque-Bera test for normality.

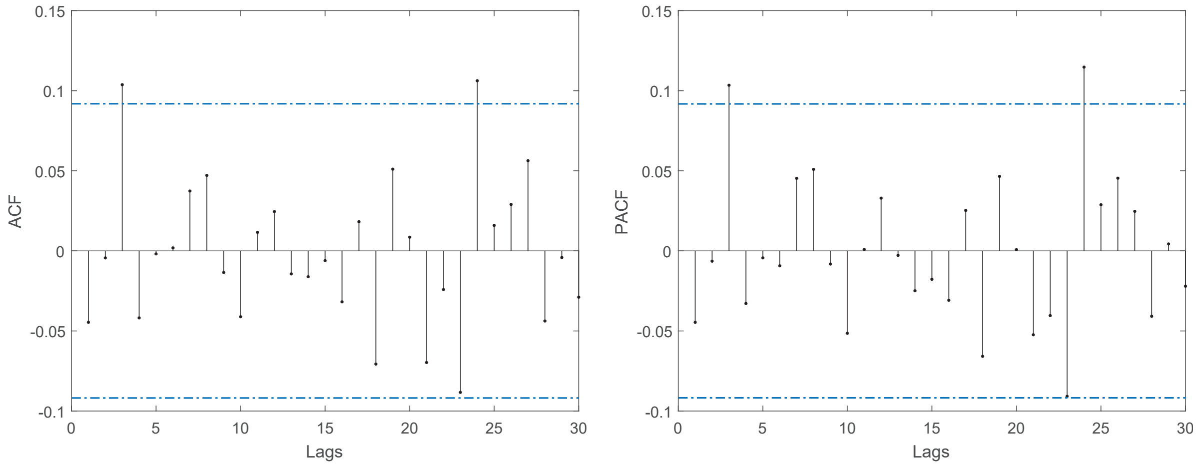
Time Period	Full sample	In-sample	Out-of-sample
Mean (%)	0.0012	0.0002	0.0023
T-stat	1.24	0.10	1.75
P-value	0.13	0.49	0.04
Minimum	-0.1945	-0.1285	-0.1945
Maximum	0.1451	0.1451	0.0912
Standard deviation	0.0311	0.0330	0.0291
Skewness	-0.24	0.16	-0.78
Z-stat	-3.06	1.47	-7.05
P-value	0.00	0.07	0.00
Excess kurtosis	2.97	2.07	4.41
Z-stat	18.87	9.22	19.95
P-value	0.00	0.00	0.00
Jarque-Bera	360.24	84.43	437.20
P-value	0.00	0.00	0.00

Table 2: Augmented Dickey-Fuller test statistic for the EUA price and log return series, testing for existence of a unit root. The considered time period spans from January 01, 2016 to December 31, 2019.

	Price time series	Log return series
Augmented Dickey-Fuller test statistic	-0.08	-22.66
P-value	0.95	0.00

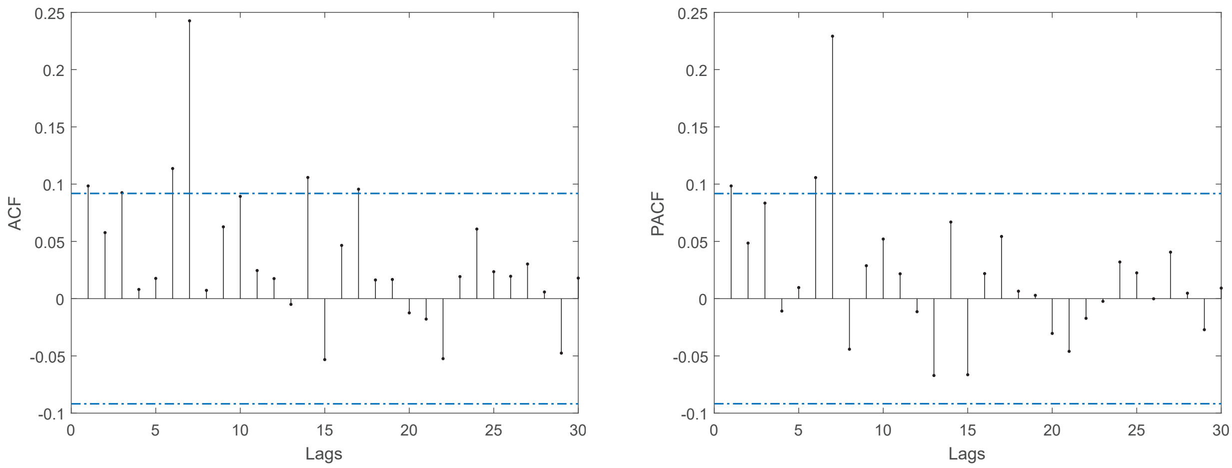
Table 3: Statistical test summary for the EUA log return and squared residual series of the ARMA(0,0) process with t-distributed error terms. $Q^2_{LB}(10)$ refers to the ten lags Ljung-Box test for serial autocorrelation.

	Log returns	Squared residuals
$Q^2_{LB}(10)$	9.60	50.88
P-value	0.47	0.00



(a) ACF **(b)** PACF

Figure 11: Sample autocorrelation function (left panel) and sample partial autocorrelation function (right panel) of EUA log returns with non-robust standard errors. The considered period is the in-sample period from January 01, 2016 to December 31, 2017.



(a) ACF **(b)** PACF

Figure 12: Sample autocorrelation function (left panel) and sample partial autocorrelation function (right panel) with non-robust standard errors of the squared residuals of an ARMA(0,0) process of the EUA daily log returns. The considered period is the in-sample period from January 01, 2016 to December 31, 2017.

Table 4: Statistical test summary for the standardized residual series of the ARMA(0,0) process with t-distributed error terms. ARCH(5) refers to Engle (1982)'s test for conditional heteroskedasticity with five lags.

ARCH(5)	4.59
P-value	0.03

Table 5: Akaike information criteria (AIC) and Bayesian information criteria (BIC) for ARMA(p,q) models with t-distributed error terms. Largest negative values for each information criterion are highlighted in bold.

	AIC				BIC			
ARMA(p,q)	0	1	2	3	0	1	2	3
0	-1925	-1924	-1922	-1924	**-1921**	-1915	-1909	-1908
1	-1924	-1922	-1922	-1923	-1915	-1909	-1905	-1902
2	-1922	-1922	**-1928**	-1926	-1909	-1905	-1907	-1901
3	-1925	-1923	-1926	-1924	-1908	-1902	-1901	-1895

Table 6: In-sample parameter estimation results for the GARCH(1,1) model with t-distributed error terms. DoF refers to degrees of freedom.

	Coefficient	Standard error	t-statistic
Mean equation			
Constant μ	0.0004	0.0013	0.33
Variance equation			
Constant α_0	0.0001	0.0000	1.50
ARCH α_1	0.1051	0.0443	2.37
GARCH β_1	0.8500	0.0614	13.84
DoF	5.9346	1.4728	4.03

Table 7: In-sample parameter estimation results for the GJR-GARCH(1,1) model with t-distributed error terms. DoF refers to degrees of freedom.

	Coefficient	Standard error	t-statistic
Mean equation			
Constant μ	0.0002	0.0013	0.12
Variance equation			
Constant α_0	0.0001	0.0000	1.60
ARCH α_1	0.0790	0.0430	1.83
GARCH β_1	0.8360	0.0657	12.72
Leverage δ_1	0.0701	0.0733	0.97
DoF	5.8633	1.4281	4.11

Table 8: In-sample parameter estimation results for the EGARCH(1,1) model with t-distributed error terms. DoF refers to degrees of freedom.

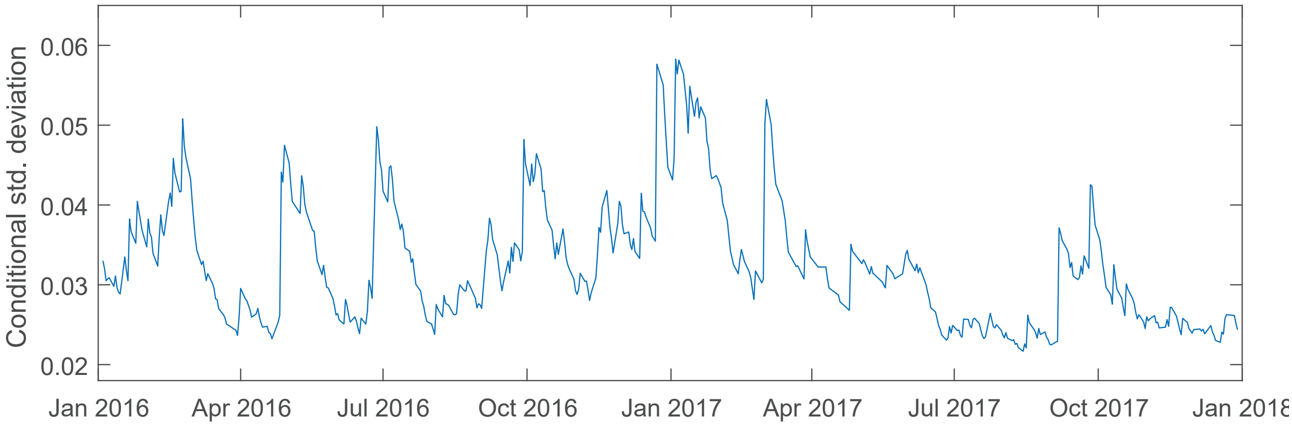
	Coefficient	Standard error	t-statistic
Mean equation			
Constant μ	0.0002	0.0013	0.17
Variance equation			
Constant α_0	-0.4151	0.2562	-1.62
ARCH α_1	0.2263	0.0804	2.82
GARCH β_1	0.9402	0.0371	25.38
Leverage δ_1	-0.0368	0.0455	-0.81
DoF	5.9398	1.5108	3.93

Table 9: Log-likelihood, Akaike information criteria (AIC) and Bayesian information criteria (BIC) for the four considered models. T-DIST refers to the simple t-distribution model, GARCH refers to the GARCH(1,1) model, GJR-GARCH refers to the GJR-GARCH(1,1) model, and EGARCH refers to the EGARCH(1,1) model.

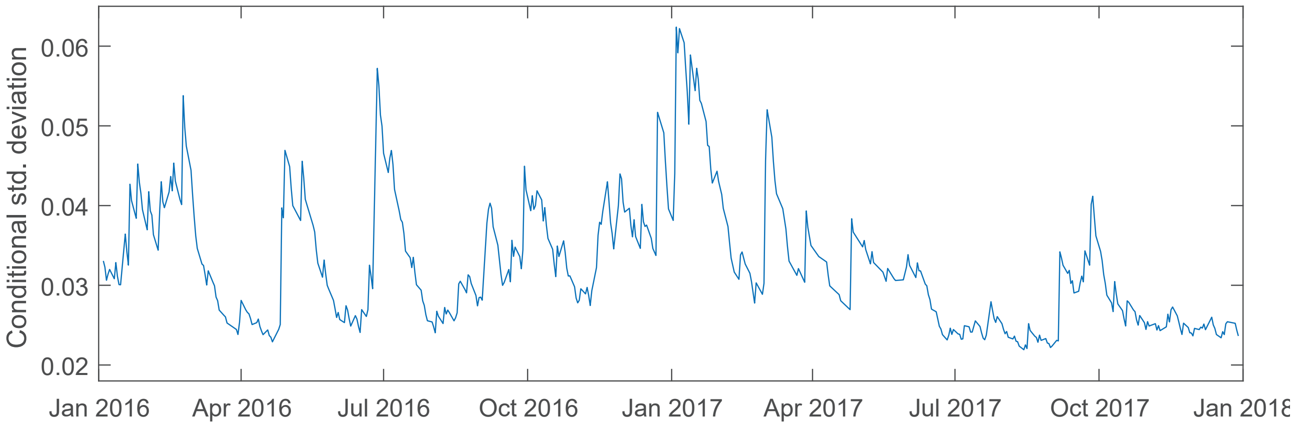
	Log-likelihood	AIC	BIC
T-DIST	963.41	-1924.82	-1920.66
GARCH	979.32	-1952.63	-1940.14
GJR-GARCH	980.08	-1952.15	-1935.50
EGARCH	980.28	-1952.55	-1935.90

Table 10: Statistical test summary for the squared residual series of the four considered models. $Q^2_{LB}(10)$ refers to the ten lags Ljung-Box test for serial autocorrelation. T-DIST refers to the simple t-distribution model, GARCH refers to the GARCH(1,1) model, GJR-GARCH refers to the GJR-GARCH(1,1) model and EGARCH refers to the EGARCH(1,1) model.

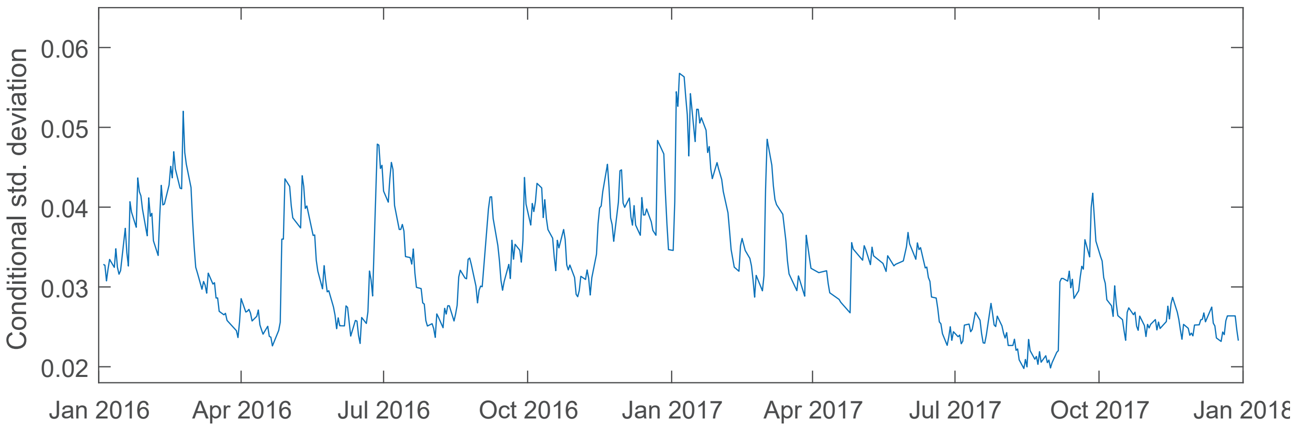
	$Q^2_{LB}(10)$	P-val
T-DIST	50.88	0.00
GARCH	6.67	0.76
GJR-GARCH	6.36	0.78
EGARCH	6.46	0.78



(a) GARCH



(b) GJR-GARCH



(c) EGARCH

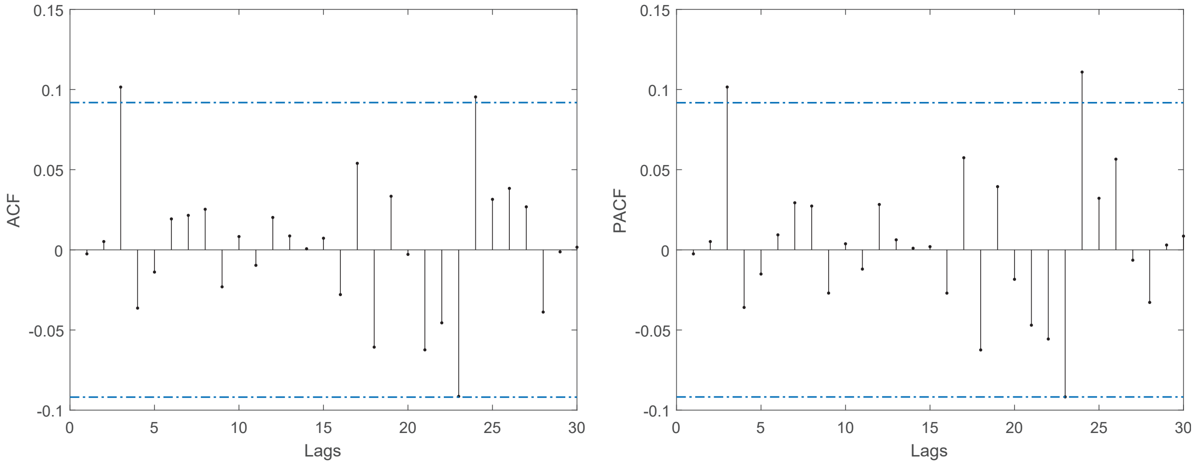
Figure 13: Inferred conditional standard deviations. Results are displayed in the respective panels for the GARCH(1,1) model (panel a), the GJR-GARCH(1,1) model (panel b), and the EGARCH(1,1) model (panel c). The considered period is the in-sample period from January 01, 2016 to December 31, 2017.

Table 11: Statistical test summary for the standardized residual series of the four considered models. ARCH(5) refers to Engle (1982)'s test for conditional heteroskedasticity with five lags. T-DIST refers to the simple t-distribution model, GARCH refers to the GARCH(1,1) model, GJR-GARCH refers to the GJR-GARCH(1,1) model, and EGARCH refers to the EGARCH(1,1) model.

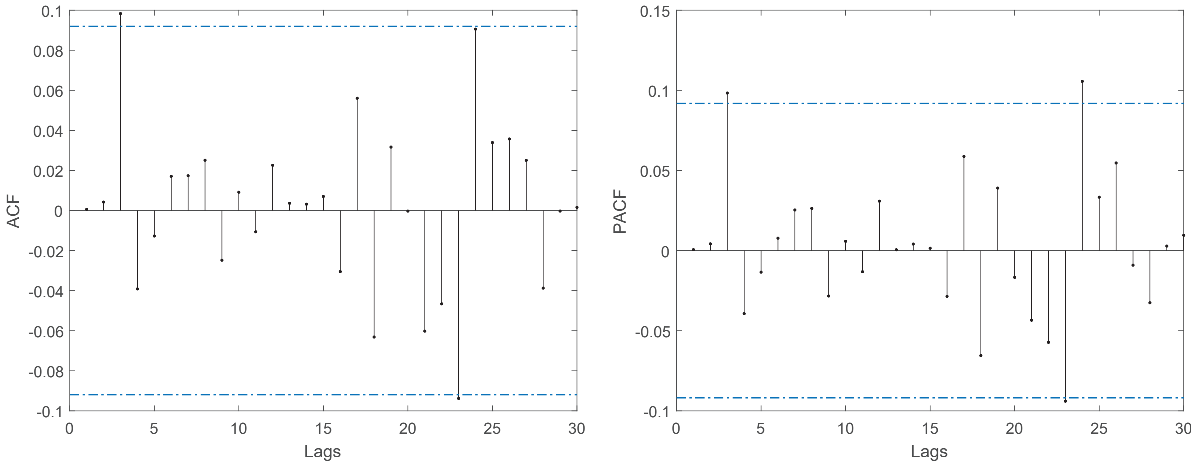
	ARCH(5)	P-val
T-DIST	4.59	0.03
GARCH	3.88	0.57
GJR-GARCH	3.35	0.65
EGARCH	3.31	0.65

Table 12: Forecasting point accuracy measures. Results for the mean-squared error (MSE) and mean absolute error (MAE) for the point forecasts of the four considered models. Smallest mean errors are highlighted in bold. T-DIST refers to the simple t-distribution model, GARCH refers to the GARCH(1,1) model, GJR-GARCH refers to the GJR-GARCH(1,1) model, and EGARCH refers to the EGARCH(1,1) model.

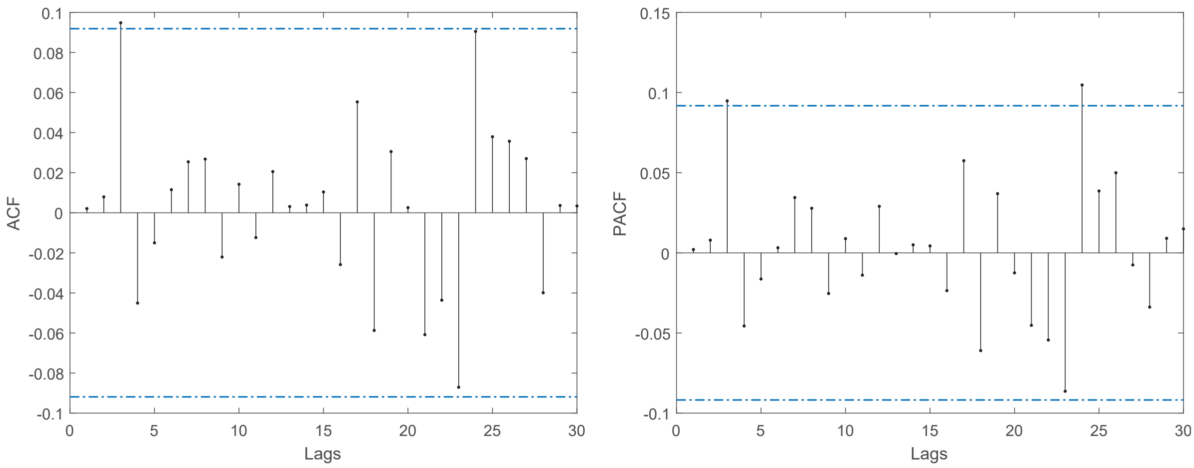
	GARCH	GJR	EGARCH	T-DIST
Rolling window				
MSE	**0.0008477**	0.0008496	0.0008494	0.0008498
MAE	**0.0216143**	0.0216288	0.0216179	0.0216601
Expanding window				
MSE	**0.0008462**	0.0008468	0.0008465	0.0008482
MAE	**0.0216403**	0.0216482	0.0216445	0.0217033
Fixed window				
MSE	**0.0008466**	0.0008477	0.0008475	0.0008505
MAE	**0.0217309**	0.0217585	0.0217518	0.0218193



(a) GARCH: ACF (left) and PACF (right)

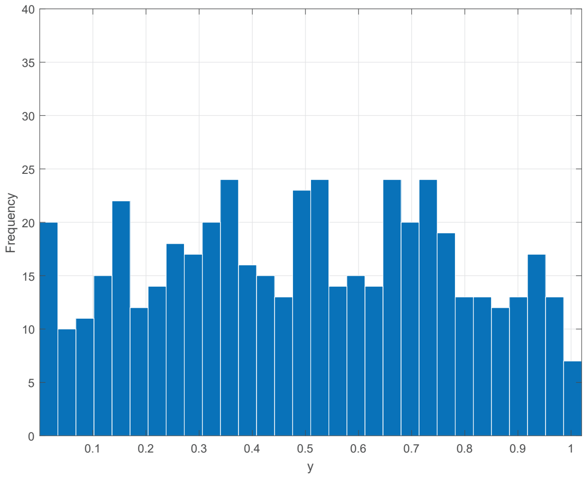


(b) GJR-GARCH: ACF (left) and PACF (right)

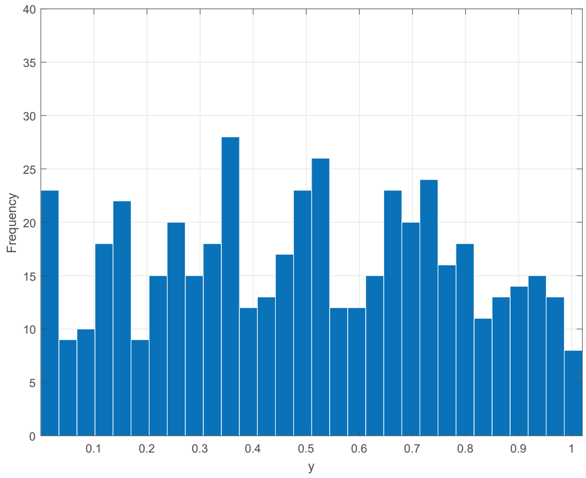


(c) EGARCH: ACF (left) and PACF (right)

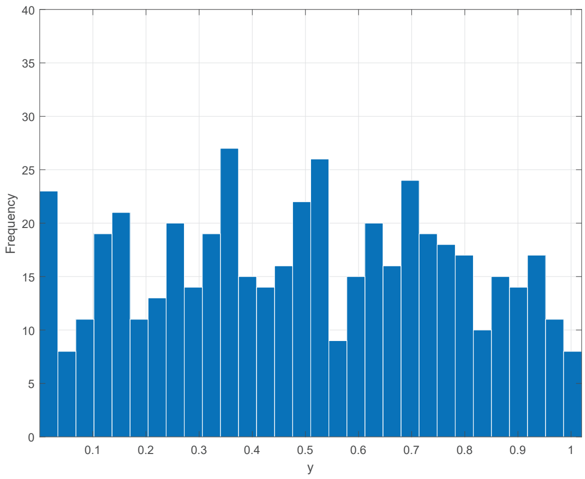
Figure 14: Sample autocorrelation function (left sub-panels) and sample partial autocorrelation function (right sub-panels) with non-robust standard errors of standardized residuals of the considered GARCH models for the EUA daily log returns in the in-sample period from January 01, 2016 to December 31, 2017. Results are displayed in the respective panels for GARCH(1,1) model (panel a), the GJR-GARCH(1,1) model (panel b), the EGARCH(1,1) model (panel c).



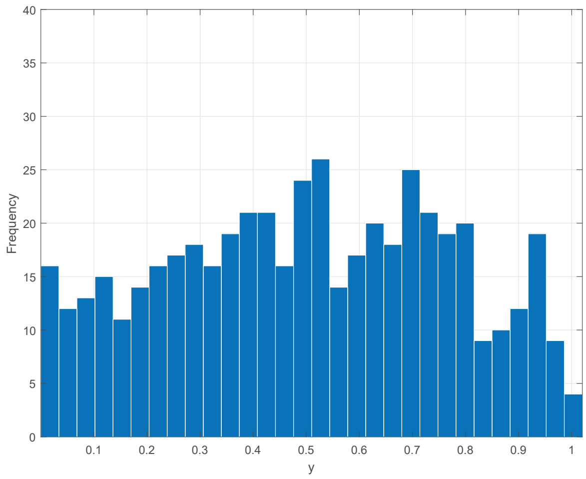
(a) GARCH



(b) GJR-GARCH

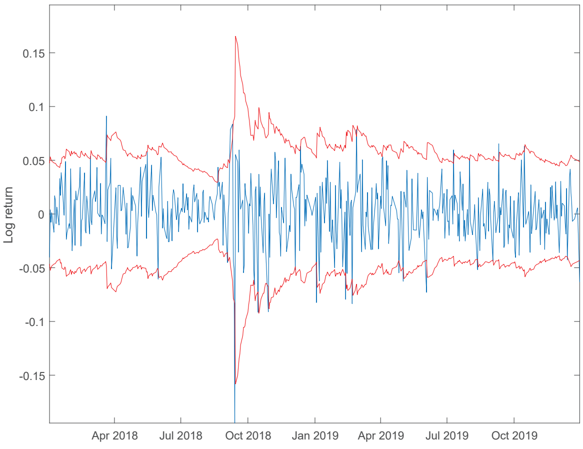


(c) EGARCH

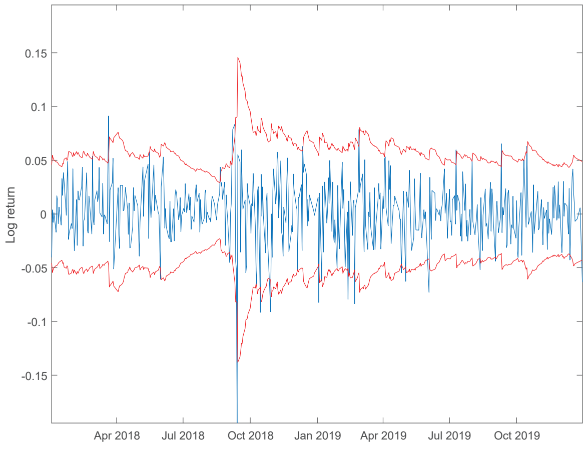


(d) T-DIST

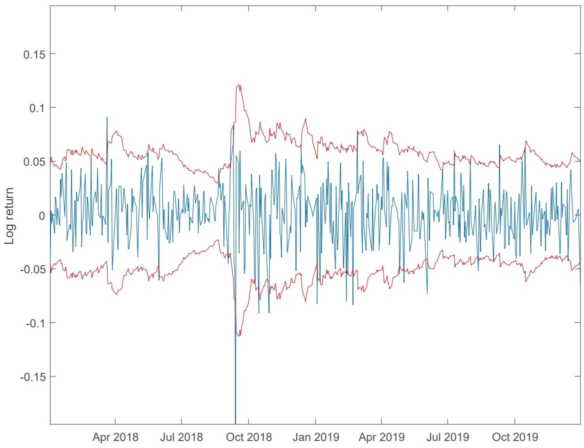
Figure 15: Histogram of the probability integral transforms of the one day ahead forecasts for the EUA daily log returns in the out-of-sample period from January 01, 2018 to December 31, 2019. Results are displayed in the respective panels for the GARCH(1,1) model (panel a), the GJR(1,1) model (panel b), the EGARCH(1,1) model (panel c), the naive model of a t-distribution (panel d). All results are based on the rolling window forecasting approach.



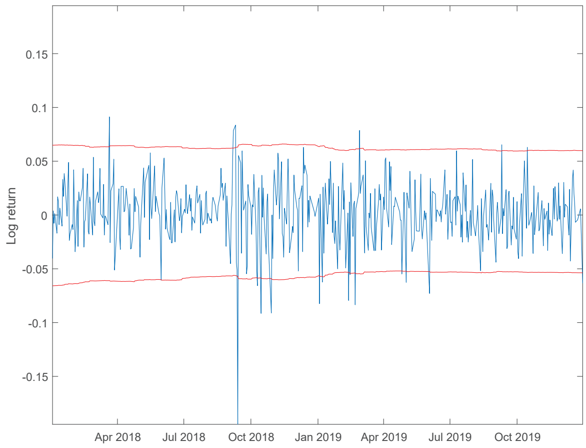
(a) GARCH



(b) GJR-GARCH

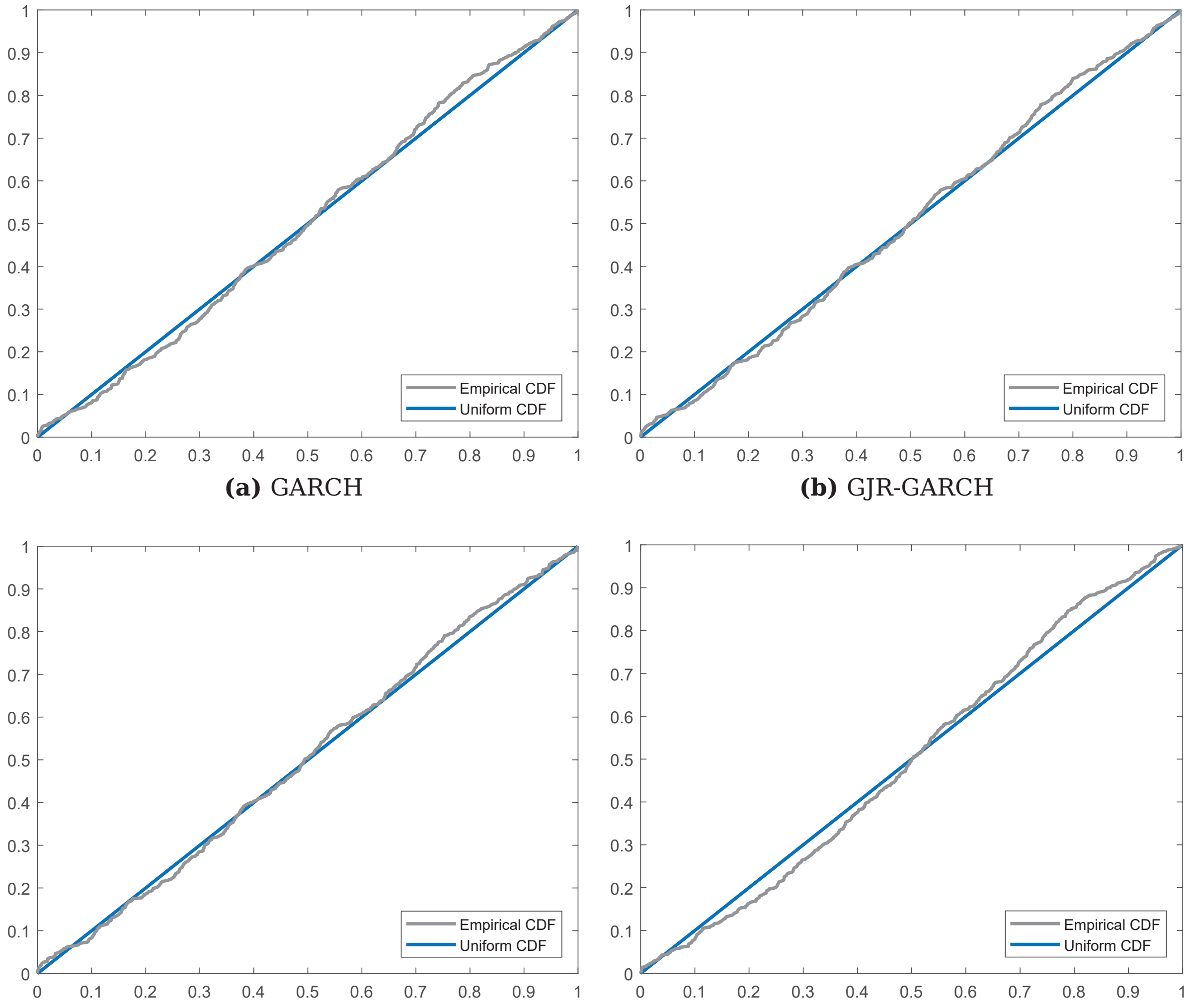


(c) EGARCH



(d) T-DIST

Figure 16: Log returns and predicted 95%-confidence intervals for the four considered models in the out-of-sample period from January 01, 2018 to December 31, 2019. Results are displayed in the respective panels for the GARCH(1,1) model (panel a), the GJR-GARCH(1,1) model (panel b), the EGARCH(1,1) model (panel c), the naive model of a t-distribution (panel d). All results are based on the rolling window forecasting approach.



(a) GARCH **(b)** GJR-GARCH **(c)** EGARCH **(d)** T-DIST

Figure 17: Cumulative distribution function of the uniform distribution and empirical cumulative distribution function of the probability integral transforms of the one day ahead forecasts of the four consider models for EUA daily log returns in the out-of-sample period from January 01, 2018 to December 31, 2019. Results are displayed in the respective panels for the GARCH(1,1) model (panel a), the GJR(1,1) model (panel b), the EGARCH(1,1) model (panel c), the naive model of a t-distribution (panel d). All results are based on the rolling window forecasting approach.

Table 13: Results for the Kolmogorov-Smirnov and Kuiper tests for the four considered models with the uniform distribution as theoretical distribution and the empirical distribution based on the respective probability integral transforms of the one day ahead forecasts. KS refers to the Kolmogorov-Smirnov test statistic. All results are based on the rolling window forecasting approach. Smallest test statistic values are highlighted in bold. T-DIST refers to the simple t-distribution model, GARCH refers to the GARCH(1,1) model, GJR-GARCH refers to the GJR-GARCH(1,1) model, and EGARCH refers to the EGARCH(1,1) model.

	KS	P-value	Kuiper	P-value
GARCH	0.0421	0.3379	0.0765	0.0615
GJR-GARCH	0.0399	0.4044	0.0652	0.2137
EGARCH	**0.0378**	0.4727	**0.0648**	0.2234
T-DIST	0.0574	0.0749	0.1062	0.0006

PART II

Emission Allowances And Natural Gas: A Cointegration Analysis

Franziska Manke

June 24, 2020

Abstract

This thesis analyses the long-run relationship between emission allowances (EUA) and European natural gas (TTF). I determine whether emission allowances and natural gas have a long-run relationship, whether this relationship has changed over the considered period, and which market leads the price discovery process. The hypothesis I suggest in support of the existence of a long-run relationship between the EUA and TTF time series is that fuel switching from coal to natural gas is one of the main practically achievable and available options to power producers in order to abate emissions and that the fuel switching decision is connected to the price of emission allowances. This thesis's empirical results suggest that there is a cointegration relationship between the two time series in the sample period covering the calendar year 2019 while showing no cointegration relation in earlier sample periods. As a logical consequence, one can assume that the relationship between the two variables has changed over time. The estimated parameter coefficients of the Vector Error-Correction Model (VECM) are employed to analyze the price discovery process. The results do not unequivocally support the assumption that EUAs drive the price discovery process.

Contents

1 Introduction

The European Union (EU) has set out the ambitious target to achieve a climate-neutral economy and society by 2050.[1] The EU Emission Trading System (EU ETS) is the primary tool of the EU to reduce greenhouse gas emissions and is designed to do so in the most cost-effective and economically efficient manner European Union (2003). In order to achieve the long term goal of net-zero emissions, new technological innovations are required. In the interim, existing solutions towards lower emissions are a vital stepping stone. The two central angles to achieve decarbonization are conservation and sequestration. Conservation technologies and processes reduce gross emitted emissions, and sequestration technologies and processes reduce net emissions. The latter refers to both natural approaches, e.g., reforestation, and carbon capture and storage technologies. While sequestration is vital in achieving a net-zero carbon goal, existing methods and technologies have not reached commercial scaleability that compares to the advances and already available solutions in conservation.

The most notable and intuitive conservation approach is the increased use of renewable energy sources in power generation. As highlighted in the first part of this thesis, the share of renewable energy has significantly increased in the European Union over the past decade. As there is inherent variability of energy production from renewable sources based on weather conditions and lacking storage technology able to handle this variability, reliable power generation still requires fossil fuels at this stage. The electricity demand and supply in a system need to balance at all points in time due to the lacking ability to store electricity effectively. Therefore, the constantly changing demand driven by economic activity is met by power production in order of cost of production, referred to as the merit order. While the merit order differs in detail across countries based on the local energy mix and natural resource availability, broadly speaking, renewable plants and nuclear plants form the basis with the lowest costs, and fossil fuel-based power generation follows with higher costs associated with production. Besides regional differences, the merit order among fossil fuels is additionally dependent on the costs arising

[1] The European Green Deal: Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the regions (COM/2019/640 final)

from emitting emissions that are captured by the EU ETS
Although the power generation sector is one of the main emissions emitters under the EU ETS, it also offers an additional source of emissions conservation that is already available, namely fuel switching from high carbon-intensive fuel sources to low carbon-intensive fuel sources. The most widely considered and relevant fuel switching pair is the switch from coal to gas. The incentive for power generators, i.e., mainly utilities, to switch from the cheaper but more carbon-intensive coal to the less carbon-intensive gas as fuel for power generation must be high enough that the related abatement costs are outweighed by the benefit of not emitting emissions and hence not having to buy emission allowances. This incentive to switch comes in the form of relative costs of the two fuel options in question and EUA prices that constitute a saving for the power producers when using gas compared to coal. While the historical price and regulatory developments of the EUA market have been outlined in detail in the first part of this thesis, the second part of the thesis revolves around the relationship between emission allowance prices and natural gas prices.

In detail, the focus of this part of the thesis is to determine whether emission allowances and natural gas have a long-run relationship, whether this relationship has changed over the considered period, and which market leads the price discovery process. The hypothesis of this thesis underpinning a significant long-run relationship between emission allowances and natural gas is that fuel switching from coal to natural gas is one of the main practically achievable and available options to power producers in order to abate emissions and that the fuel switching decision is connected to the price of emission allowances. This hypothesis also incorporates the assumption that the emission allowances market is driving the price discovery and therefore is impacting the demand for natural gas versus the alternative assumption that natural gas prices are driving the price discovery as the lower emissions emitted from switching fuel from coal to gas is impacting the demand for emission allowances. Independent on whether the hypothesis or alternative hypothesis of the price discovery is supported by empirical evidence, the relationship of emission allowances and natural gas is assumed to be positive. The analysis of this thesis extends the existing literature by using the latest market data and hence capturing the latest regulatory and market developments. This thesis's empirical results suggest that there is a cointegration relationship between the two time series and that this relationship has only developed in 2019 compared to non-significant cointegration results for earlier sub-samples starting from 2014. The price discovery measures indicate that emission allowances are incorporating information first. However, the empirical results do not appear to be unequivocally clear. The following sections outline the relevant literature, methodology, and empirical results.

2 Literature Review

In the initial phase of emission trading, researchers focused on finding factors influencing the behavior of EUAs. Christiansen et al. (2005) identify three main factors impacting EUAs: policy and regulatory factors, market fundamentals, and technical indicators. Analyzing data from the first year of the first phase of the EU ETS, Mansanet-Bataller et al. (2007) show that energy sources principal factors in the determination of EUA prices and find that oil prices, natural gas prices but also temperatures in Germany are significant EUA price drivers. Alberola et al. (2008) confirms the findings of Mansanet-Bataller et al. (2007) and additionally shows that there are structural breaks in the emissions time series in the first phase of the EU ETS. In one of the first papers incorporating the cointegration relationship of energy markets and the EUA market, Bunn and Fezzi (2007) show EUA prices are cointegrated with British naturals gas (NBP) prices and UK electricity prices. Using a cointegrated VAR model, the authors show that gas prices are driving carbon prices, and both gas and carbon prices are driving electricity prices. In contrast, Rickels et al. (2007) do not find a cointegration relationship among EUAs, oil, natural gas, and coal prices using data from the first phase of the EU ETS. Compared to Bunn and Fezzi (2007) and besides the focus on various input prices for power generation, the authors use Zeebruegge natural gas (ZTP) prices rather than the NBP price series used by Bunn and Fezzi (2007). It is further noteworthy that Rickels et al. (2007) find a cointegration relationship under the Johansen approach, but testing the variables shows that oil, natural gas, and coal are weakly exogenous to the error correction term and that the application of a single equation error correction model yields strong evidence against cointegration. In a later study using data from both the first and second phase of the ETS, Bredin and Muckley (2011) show that the cointegration relationship between emissions futures and fundamentals, including economic growth, weather, energy spreads (clean dark and spark spreads for coal and gas to electricity, respectively), and oil prices, is more evident in the second phase of the EU ETS. The authors employ three different cointegration frameworks, the Engle-Granger approach, the Johansen approach, and a modified cointegration approach accounting for heteroskedasticity and analyze the development over time. While the Engle-Granger approach does not show a significant cointegration relationship throughout the two phases, both the Johansen and modified cointegration approach show that the cointegration is developing over the second phase. Bredin and Muckley (2011) conclude that the second phase of the EU ETS is more efficient than the pilot phase with higher trading volumes and that theoretically established relations are empirically apparent. Creti et al. (2012) builds upon the research of Bredin and Muckley (2011) and shows that there exists a cointegration relationship between EUAs and fundamentals for both the first and second phase of the EU ETS. The fundamentals used in this study are an equity index, oil prices, and switching costs between gas (NBP prices) and coal (API2 - Rotterdam coal futures). The positive cointegration results for the first phase are dependent on accounting for the structural break in 2006. Creti et al. (2012) additionally conclude that the cointegration relationship shifts towards a fundamentals-driven relationship in the second phase of the EU ETS and that the switching costs were not a significant driver of the long-run relationship in the first phase. While the literature extends in findings of EUA market drivers in a short term setting, the above literature review represents the core of literature focused on the long-run relationship of emissions prices and energy prices, including natural gas.

3 Methodology

This section provides an overview of methodologies relevant for the analysis of the relationship between the daily emission allowances and the natural gas price time series. In a first step, the univariate analysis methodologies, namely unit root testing and stationarity testing, are outlined as the outcome of such analysis forms the basis and is prerequisite for the following multivariate time series analysis. In the multivariate space, the concepts of and approaches to cointegration analysis are forming the core of the Methodology section and this thesis. While two prominent methodologies of cointegration analysis are presented, the focus is set on the Johansen approach (Johansen, 1988) as this is the main framework applied in this thesis.

3.1 Unit Roots and Stationarity

Similar to the univariate analysis of the emission allowances returns in Part I of this thesis, the concept of stationarity forms a crucial role in the context of multivariate analysis and specifically, cointegration analysis. Compared to the analysis discussed in Part I, the focal point of this Part of the thesis is the price time series, i.e., levels of the data, rather than the log returns, i.e., differences of the data. Although the prices of assets in the financial context are generally considered non-stationary, a more refined analysis is required to formulate the appropriate model and derive meaningful conclusions from the cointegration analysis. Conversely to the analysis done in Part I, the implicit goal, i.e., fulfilling prerequisites for subsequent analysis, of the unit root and stationarity testing is not to show that the time series are stationary but rather that the time series under consideration are integrated to the same order and non-stationary in levels. More details regarding cointegration analysis prerequisites are provided in the respective section.

As mentioned in Part I of this thesis and as Brooks (2002) outlines, non-stationary time series need to be treated differently from stationary ones as the non-stationarity can strongly influence a time series's behavior and properties, can lead to spurious regressions, and a regression's standard assumptions for asymptotic analysis are not valid. The most restrictive form of stationarity is the strict stationarity, which requires a stochastic process, y_t, to have a joint distribution of $\{y_{t,}, y_{t+1}, \ldots, y_{t+h}\}$ to depend only on h and not t. This assumption is generally not fulfilled by financial time series as it implies the distribution is independent of time. In contrast, a weakly or also covariance stationary process requires the series to have a constant mean, constant variance, and constant autocovariance. Generally, financial return series, such as the emissions log return series discussed in Part I of this thesis, are expected to be weakly stationary.

In the realm of non-stationary time series, two types can be frequently observed; the stochastic and deterministic non-stationarity. The two forms are expressed in Equation 1 and Equation 2, respectively

$$y_t = \mu + y_{t-1} + u_t \tag{1}$$

$$y_t = \alpha + \beta t + u_t \tag{2}$$

Stochastic non-stationary processes are also called unit root processes, random walks with drift, and difference stationary processes. The latter alludes to the fact that the time series is stationary in differences; hence, the process described by Equation 1 takes the following form in differences.

$$\Delta y_t = \mu + u_t \tag{3}$$

The stationarity of the first differences implies the time series y_t of Equation 1 is integrated of order $I(1)$. More generally, a time series is said to be integrated of order $I(d)$, where d is the number of times the series is differenced. Hence a stationary process is integrated of order $I(0)$ as it is already stationary itself.

In contrast, deterministic processes are not stationary in differences but require detrending to derive a stationary series. A deterministic non-stationary process is also called a trend stationary process as the series is stationary around the trend.

To identify the correct underlying data generating process, one shall firstly plot the time series and the first differences. Additional graphical support can be provided by an autocorrelation function. While experienced researchers may have developed a good sense from purely inspecting the graphical evidence, a more formal procedure is required to confirm the initial prognosis. Common tests for this purpose are the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979), the Phillips-Perron (PP) test (Phillips and Perron, 1988), and the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test (Kwiatkowski et al., 1992). Importantly, the ADF test and PP test are testing for the existence of a unit root in the process, while the KPSS test is testing for the stationarity of it.

The ADF test allows for different variations, and its most basic form has been discussed in Part I of this thesis. The full spectrum of the ADF test's regression set up is described by

$$y_t = c + \delta t + \phi y_{t-1} + \beta_1 \Delta y_{t-1} + ... + \beta_p \Delta y_{t-p} + \varepsilon_t \tag{4}$$

where Δ is the differencing operator, p is the number of lagged difference terms, β_p are the difference coefficients, c is the drift coefficient, δ is the deterministic trend coefficient, ϕ is the AR(1) coefficient, and ε_t is the innovation process which is assumed to be a white noise process. The null hypothesis is that $\phi = 1$,i.e., the data contains a unit root, against the alternative $\phi < 1$. The ADF test produces test statistics that do not follow a standard distribution as the null hypothesis is that the process is non-stationary. The critical values are derived by simulation. The researcher must choose the type of model and the lag length p for the ADF test. The model type specifies whether the drift coefficient c and the deterministic trend coefficient δ is present in the alternative hypothesis. The choice is depending on the underlying data generating process and can be assessed via t-test and F-test of the coefficients in questions. The plot of the time series helps determine what model choice is likely to be appropriate, but the formal assessment of t-stat and F-stat will provide more clarity. The lag length p is another input to be determined by the researcher and one that the test outcome is sensitive to. Information criteria can be used in order to find the best model fit and determine the lag length. Popular information criteria (IC) are Akaike's IC (AIC) (Akaike, 1974) and the Bayesian IC (BIC) (Schwarz, 1978). The different criteria may select the same or different lag length as they penalize additional parameters to varying degrees.

The Phillips-Perron test uses a similar set up to the ADF test, and the regression equation is

$$y_t = c + \delta t + \alpha y + e(t) \tag{5}$$

where c is the drift coefficient, δ is the deterministic trend coefficient, α is the AR(1) coefficient, and $e(t)$ is the innovation process. The test statistics are modified Dickey-Fuller stats to account for the serial autocorrelation of the innovation process $e(t)$ as the regression formula does not include lagged difference terms that account for the autocorrelation in the ADF test. The null hypothesis is that $\alpha = 1$,i.e., the data contains a unit root, against the alternative $\alpha < 1$. The lag length and model selection procedure is equivalent to the one used in the context of the ADF test.

Compared to the ADF and PP tests, the KPSS test has a null hypothesis that the y_t is stationary rather than testing the existence of a unit root, which implies non-stationarity. In that sense, the KPSS test is confirmatory to the ADF and PP test, and the combined conclusions from the tests should match so that if the null under either the ADF or PP test is rejected, the null should not be rejected under the KPSS test and vice versa. This is particularly useful as the ADF and PP tests have a weakness in that they tend to have relatively low power and may not reject the null because the null is true but because there is insufficient evidence in the data to suggest it was not. The KPSS's test null hypothesis is stationarity, so the default is the opposite. Kwiatkowski et al. (1992) define a structural model

$$y_t = c_t + \delta t + u_{1t} \tag{6}$$

$$c_t = c_{t-1} + u_{2t} \tag{7}$$

where δ is the trend coefficient, u_{1t} is a stationary process, u_{2t} is an i.i.d. with zero mean and variance σ^2, and c_t is the random walk term. The null hypothesis is that $\sigma^2 = 0$, implying the random walk term c_t would act as an intercept and hence y_t was stationary. In contrast, if $\sigma^2 > 0$, under the alternative hypothesis, a unit root is introduced in the system. Similarly to the ADF and PP tests, the KPSS test can be performed with or without the deterministic trend term δ depending on the underlying data generating process. For the lag length Kwiatkowski et al. (1992) suggest $\sqrt{T}$, where T is the sample size, is an appropriate lag length.

The tests discussed in this subsection help establish whether a univariate time series is stationary or non-stationary and help determine the time series's order of integration. This information is the basis for the applicability of the subsequent multivariate analysis.

3.2 Cointegration

Cointegration refers to the relationship between two or more time series and their tendency to move together over time. This concept is commonly observed in various markets and grounded in economic theory. Examples include spot and futures markets of the same assets, the same asset traded in different locations, and substitute assets. Engle and Granger (1987) first formally defined the cointegration concept. According to Engle and Granger (1987) the components of a vector y_t are cointegrated of order $CI(b, d)$ if all components of x_t are $I(d)$ and there exists a vector $\alpha \neq 0$ so that

$$z_t = \alpha' x_t \sim I(d - b) \tag{8}$$

with $b > 0$. In other words, the components of x_t are cointegrated if there is a stationary linear combination of the components. Most financial time series are $b = d = 1$, implying the innovation process is stationary, i.e., $I(0)$. Cointegrated variables have a long-term relationship but may deviate from that relationship in the short term. The dynamics of the deviation and reversion from and to the long term relationship are captured in error-correction models. Economic theories provide the initial hypothesis for long term relationships of economic and financial variables. However, empirical methods are required to confirm such theories. In the following two main testing and estimation approaches are presented; the Engle-Granger approach (Engle and Granger, 1987) and the Johansen approach (Johansen, 1988). This thesis applies and hence focuses on Johansen's approach.

3.2.1 Engle-Granger approach

The Granger representation theorem Engle and Granger (1987) states that if there is an error correction model, i.e., a dynamic linear model with $I(0)$ innovations, exists, and the underlying data components are $I(1)$ the components are co-integrated. The Engle-Granger or residual based approach makes use of this fact employs an OLS regression of one variable on the other components and tests the residuals for stationarity. The approach assumes the pre-testing of individual variables for integration such that all variables are $I(d)$ for $d > 0$. Considering the general regression

$$y_t = \beta_1 + \beta_2 x_{2t} + ... + \beta_k x_{kt} + u_t \tag{9}$$

where u_t represents the innovation process that is stationary if y_t,x_{2t},...,x_{kt} are cointegrated. While the ADF test approach to unit root testing can be applied, the critical values of the test distribution are not equal to the one in the standard application as the innovation process represents an estimated series itself and not raw data. Engle and Yoo (1987) develop a new set of critical values to account for this fact. If the null hypothesis of $\hat{u}_t \sim I(1)$ is rejected and the alternative hypothesis of $\hat{u}_t \sim I(0)$ consequently accepted the variables y_t,x_{2t},...,$x_{kt}$are cointegrated. The Engle-Granger approach can at most detect one cointegration relationship while there may be more relationships among the variables. However, if only two variables are included, there can only exist one cointegration relationship by definition, and this approach restriction does not cause issues. If the testing has concluded there is a cointegration relationship, the parameters of the error-correction model can be estimated in a second step. In this second step, the residuals are inserted into the error-correction model such that the model is defined as

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2(\hat{u}_{t-1}) + v_t \tag{10}$$

where $\hat{u}_{t-1} = y_{t-1} - \hat{\tau} x_{t-1}$ and since $\hat{u}$ is stationary as tested in the first step, any linear combination is equally stationary, and OLS is an appropriate approach for model estimation. It is noteworthy that the Engle-Granger approach requires the researcher to specify one variable, here y_t, as the dependent one while this selection may not represent the actual or theoretical relationship. In addition, the Engle-Granger approach does not allow for hypotheses testing. However, the Johansen approach overcomes these shortcomings.

3.2.2 Johansen approach

The Johansen approach is based on the Vector-Autoregressive (VAR) model, which corresponds to the univariate case of AR processes discussed in Part I of this thesis. The

VAR model is augmented by inclusion of the cointegration term resulting in the Vector Error-Correction (VEC) model as describes by

$$\Delta y_t = C y_{t-1} + \sum_{i=1}^{q} B_i \Delta y_{t-i} + \varepsilon_t \tag{11}$$

where Cy_t is the error-correct term which itself is described by Equation 12. The remaining part of Equation 11 is equal to the VAR model. The error-correction term is described by

$$C y_{t-1} = A B' y_{t-1} \tag{12}$$

where B is the cointegrating matrix that measures the error, meaning the deviation from the equilibrium, and A represents the adjustment speeds or the rate at which the process returns to equilibrium after a given shock. The number of lags q in the VEC(q) model is related to the lag order of a VAR(p) model such that $q = p - 1$. The lag length is impacting the Johansen test outcome, and hence the lag selection is a vital step that can be accomplished via information criteria. The parameter estimation is done via the maximum likelihood approach.

The Johansen test is determining whether a process is cointegrated based on the rank of C given that in equilibrium Equation 11 reduces to

$$C y_{t-1} = 0 \tag{13}$$

as $\sum_{i=1}^{q} B_i \Delta y_{t-i}$ and $E[\varepsilon_t]$ are equal to zero in equilibrium. Therefore if there was a cointegration relationship, C must have a rank of larger than zero so that there is a non-zero solution to Equation 13. If there was no non-zero solution, i.e., rank = 0, Equation 11 reduces to the VAR model in first differences. If matrix C has full rank, it is implied that y_t is stationary. So if there are one or more cointegration relationships, the rank r of matrix C, with g dimensions, has to fulfill $0 < r < g$.

The Johansen test employs two testing methods with test statistics presented in Equation 14 and 15 for the trace test and the maximum eigenvalue test, respectively.

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{g} ln(1 - \hat{\lambda}_i) \tag{14}$$

$$\lambda_{max}(r, r+1) = -T\ ln(1 - \hat{\lambda}_i) \tag{15}$$

Where $\hat{\lambda}_i$ are the eigenvalues that are sorted in ascending order. The trace test is a joint test with the null hypothesis that the number of cointegrating vectors is less than or equal to r. The maximum eigenvalue test is testing each eigenvalue separately whether the eigenvalue is equal to r. Both test methods are done in a sequential manner so there is a test statistic for each r . The test statistics do not follow a standard distribution, and critical values are provided by Johansen and Juselius (1990). The Johansen testing set up additionally allows for hypothesis testing within the cointegration matrix C that consists of matrices A and B, as described in Equation 12. Restrictions on A or B can be imposed by substituting them into the respective matrices. If the eigenvalues under the restricted model are not significantly different from the unrestricted model the test statistic in Equation 16 shall not exceed $\chi^2(m)$ distributed critical values with m number of restrictions.

$$test\ statistic = -T \sum_{i=1}^{r} [ln(1 - \lambda_i) - ln(1 - \lambda_i^*)] \sim \chi^2(m) \tag{16}$$

Similarly to the univariate case discussed in the context of testing for a unit root or stationarity, deterministic terms in the data generating processes are essential to take into consideration to correctly specify models and be able to draw accurate conclusions from testing procedures. Incorporating deterministic terms in Equation 11 it becomes

$$\Delta y_t = C y_{t-1} + \sum_{i=1}^{q} B_i \Delta y_{t-i} + Dx + \varepsilon_t \tag{17}$$

where Dx is an exogenous term that may represent constants, linear deterministic trends, or quadratic deterministic trends in the levels, i.e. y_t, of the data. Johansen (1995) identifies five forms that $Cy_{t-1} + Dx$ can take covering the majority of applications in financial and economic time series. The first and most basic form is described by Equation 12, implying there are no intercepts or trends in the cointegrated series and no intercepts or deterministic trends in the levels of the data. The second form, as shown in Equation 18, additionally includes an intercept in the cointegrated series but no deterministic terms in the levels of the data.

$$Cy_{t-1} + Dx = A(B'y_{t-1} + c_0) \tag{18}$$

The third form adds a constant c_1 outside of the cointegrated series to account for a deterministic trend in the levels of the data. While this term is expressed as a constant in the Equation 19, it appears as a trend in the levels as the VEC model is expressed in differences.

$$Cy_{t-1} + Dx = A(B'y_{t-1} + c_0) + c_1 \tag{19}$$

The fourth form adds a deterministic trend to the cointegrated series resulting in Equation 20.

$$Cy_{t-1} + Dx = A(B'y_{t-1} + c_0 + d_c t) + c_1 \tag{20}$$

The last and fifth form incorporates $d_1 t$ outside of the cointegrated series and hence accounting for quadratic deterministic trends in the levels of the data.

$$Cy_{t-1} + Dx = A(B'y_{t-1} + c_0 + d_c t) + c_1 + d_1 t \tag{21}$$

The appropriateness of the model choices can be confirmed via pairwise likelihood ratio tests with the test statistic

$$test\ statistic_{LR} = 2 * [uLL - rLL] \sim \chi^2(dof) \tag{22}$$

where uLL is the unrestricted likelihood of the unrestricted model, rLL is the restricted likelihood of the restricted model, and dof refers to the degrees of freedom that are equal to the number of restrictions imposed.

3.2.3 Price Discovery

Price discovery measurement is a vital tool to bridge theoretical assumptions of the relationship of markets with the empirical results. Price discovery refers to the concept that one market will reflect new information more timely than the others and will drive the formation of prices in those markets. Two popular measurements are the component share (CS) based on Gonzalo and Granger (1995) and the information share (IS) based on Hasbrouck (1995). Both measures are common factors models.

The CS measure or common factor weight is captured by

$$CFW_a = \frac{-\delta_b}{\delta_a - \delta_b} \tag{23}$$

where δ_a and δ_b are the adjustment speeds for the respective markets. This measure is attributing price leadership to the market that has the lowest rate of adjustments or in other words adjusts the least to price information in the other market. One can observe that if $\delta_a = 0$, Equation 23 reduces to 1, implying market a is solely leading the price discovery.

The IS measure consist of a lower and upper bound captured by

$$IS^a_{lower} = \frac{\left(\delta^b \sigma^a - \delta^a \frac{\sigma^{a,b}}{\sigma^a}\right)^2}{(\delta^b \sigma^a)^2 - 2\delta^a \delta^b \sigma^{a,b} + (\delta^a \sigma^b)^2} \tag{24}$$

and

$$IS^a_{upper} = \frac{(\delta^b)^2 \left((\sigma^a)^2 - \frac{(\sigma^{a,b})^2}{(\sigma^a)^2}\right)}{(\delta^b \sigma^a)^2 - 2\delta^a \delta^b \sigma^{a,b} + (\delta^a \sigma^b)^2} \tag{25}$$

where δ_a and δ_b are the adjustment speeds for the respective market and $(\sigma^a)^2$,$\sigma^{a,b}$, and $(\sigma^b)^2$are the constituents of the covariance matrix of the innovations of the respective VEC model (representative of ε_t in Equation 11). Compared to the CS measure, the IS measure is attributing price leadership to the market for which the contribution to the variance of the innovations is the highest. Baillie et al. (2002) highlight that if the innovations ε_t are uncorrelated, the two measures, CS and IS, are related. However, if the correlation of ε_t is high, the concepts are not related, and the spread between the upper and lower bounds of the IS measure is wide. Baillie et al. (2002) suggest the average of the two bounds to be a sensible measure.

4 Empirical results

4.1 Data

The data for this thesis was collected via Bloomberg. It consists of the daily EUA spot price data from the European Energy Exchange (EEX) and the daily 1-day ahead physical forward prices for natural gas delivered to the Virtual Trading Point Netherlands Title Transfer Facility (TTF). More details regarding the emissions data is provided in the respective section in Part I of this thesis. The TTF natural gas (in following abbreviated to TTF) prices are used in this thesis as a proxy for European natural gas prices in order to capture consumers of European natural gas, including power generation assets, that are also under the scope of the EU ETS. The TTF trading point is one of the main continental European hubs, and hence it is a reasonable proxy for the purpose of this thesis. The TTF prices are expressed in EUR/Mwh. As a first step in the preparation of the data, the trading days are matched. All trading days that do not have an entry for both time series are removed. Due to extreme volatility in the TTF prices in March 2018, two outliers, which are a multiple of the average price, are removed. The full considered period spans from January 01, 2014 to December 31, 2019 (Full Sample in the following). Therefore the entire time period falls within the same EUA trading phase avoiding major structural regulatory breaks. In order to determine whether the relationship between EUAs and TTF prices have changed over time, three equally sized sub-periods are considered. The sub-periods are reflecting two calendar years each such that the first sub-period spans from January 01, 2014 to December 31, 2015 (Sample 1 in the following) , the second from January 01, 2016 to December 31, 2017 (Sample 2 in the following), and the third January 01, 2018 to December 31, 2019 (Sample 3 in the following). In addition to the three equally sized sub-periods, a fourth sub-period (Sample 4 in the following) of the most recent full year, i.e., 2019, is considered. This additional sub-period is considered in order to assess whether the recently increased momentum in the carbon market reflected in the consistently high price levels during 2019 is changing the dynamics of the long-run relationship of natural gas and emissions prices.

The price time series and the price histograms of the Full Sample depicted in Figure 1 and Figures 2 and 3, respectively, clearly indicate that the EUA prices have seen a shift towards a higher price level while the TTF prices do not appear to indicate a structural shift sustained over time. The descriptive statistics of the time series for the full and sub-periods are shown in Table 1. The statistics confirm the assertion that the mean of the EUA price series is increasing over time and that the TTF price does not indicate such behavior. Additionally, the statistics highlight the relatively higher standard deviation of both the EUA and TTF prices from 2018 going forward. Noteworthy is also the very low standard deviation in sub-sample 1 and 2 for the EUA price series. In aggregation, if one were to assume there was a relationship between the two time series, it appears likely that this relationship would have changed over time. However, from this visual inspection, it is not clear if there is a relationship and the nature of that relationship.

The following empirical results subsections establish whether there is a relationship, it has changed over time, and what is the directionality of the relationship. The following subsections are organized into three components that are building upon the results of the respective previous sections. In a first step the univariate variables are pre-tested for unit roots and stationarity. The following subsection is presenting the results of the cointegration analysis itself and the tests for different model specifications. In the last subsection of the empirical results, the price discovery measures for established cointegration relationships are presented.

4.2 Unit Root and Stationarity Testing

Establishing the integration $I(d)$ for each of the considered variables is vital to conduct the cointegration analysis. All variables, i.e., the EUA and TTF time series, need to be integrated of order d. As both time series are price time series, it is reasonable to suspect the order of integration is one, meaning that the levels of the data are non-stationary while the first differences, the returns, are stationary. This initial assumption can be firmed up by plotting the time series and their first differences before concluding the analysis with formal tests. Figures 2 to 11 show the price and return series for both EUA and TTF underliers for the full and all sub-samples. The plots of both time series show standard patterns in log returns that support the assumption both time series are integrated of order one. The plots of the prices and descriptive statistics clearly show that the mean in both price time series is significantly different from zero. In addition, the mean and variance of both time series appear to change over time, as evidenced by the varying levels of mean and standard deviation presented in Table 1. However, weak stationarity requires a constant mean, variance, and autocovariance structure. The visual queues and descriptive statistics points support the notion that the levels of the data are not stationary. In order to confirm this assumption and therefore confirm the two price time series are suitable for cointegration analysis, the ADF, PP, and KPSS tests are used as formal statistical tests.

The ADF test requires lag length choice and model specification choice to reflect accurate results. In a first step, the lag length is assessed using the information criteria AIC and BIC with optimal choices reported in Table 2. Unsurprisingly, the AIC is selecting a higher lag order than the BIC for most of the sub-samples as the BIC is putting higher penalties on additional parameters. The t-test and F-test results of the ADF test for both the EUA and TTF price series are reported in Table 2 for each possible model specification at optimal lag length determined by AIC and BIC respectively. The decision outcome based on a 5% significance level is consistent across lag length choice via the two criteria despite the sometimes significantly different lag length.

In contrast, the model specification choice does influence the decision outcome for some of the sub-sample based on a 5% significance level. While the Figures 2 and 3 for the full sample period do not seem to indicate the presence of deterministic terms in the data generating process, the individual sub-samples, assessed independently from the full and other sub-samples, show signs that deterministic terms may be present. The ADF test is hence conducted for all model specifications, i.e., AR, ARD, and TS. The model specification AR refers to Equation 4 where $c = \delta = 0$. The model specification ARD refers to Equation 4 where $\delta = 0$. The model specification TS refers to Equation 4. For the ARD test, the F-test assesses the joint hypothesis that $\phi - 1 = 0$ and $c = 0$, and for the TS specification, the F-test assesses the joint hypothesis that $\phi - 1 = 0$ and $\delta = 0$. The p-values for each test and model specification at optimal lag length for AIC and BIC are presented in Table 2.

The results indicate that for EUA prices, the F-test is rejecting the joint null at the 5%

significance level for Sample 1 and Sample 2 for each lag specification. This result would imply the TS model specification, including a deterministic drift and trend term, is appropriate for Sample 1 and Sample 2 of the EUA series. However, looking at the coefficient estimate for δ it becomes clear that the magnitude is negligible with $\delta_1 = 0.0004$ and $\delta_2 = 0.0002$ for Sample 1 and Sample 2, respectively. The deterministic coefficients for all EUA and TTF samples are reported in Table 3. The EUA Sample 1 results for the TS model specification indicate that the null of a unit root is rejected at the 5% significance level. The EUA Sample 2 results can only reject the null at the 10% significance level. The other model specifications for these two samples show that the null of a unit root cannot be rejected. Based on the small magnitude of the deterministic trend term and the lacking rejection under other model specifications, it appears difficult to make a definite decision on the existence of a unit root for the EUA Sample 1 series. The EUA Sample 2 series seems to not have significant enough support even at the TS model specification to come to the conclusion that there is no unit root. The results for the TTF data series are more compellingly clear. All lag length, model specification, and sample period combinations cannot reject the null hypothesis of a unit root. Similarly to the EUA data series, the deterministic terms have little magnitude overall.
Therefore, one can conclude that, with the exception of EUA Sample 1 and Sample 2, it is reasonable to assume that all sample time series are non-stationary. The ADF test results for the log return series are not reported, but all combinations of lag length and model specification for all samples of the EUA and TTF time series are highly significant, implying the first differences of the original time series are stationary according to the ADF test. The combination of both of these findings implies the price time series are integrated of order one, i.e., $I(1)$. As there are some doubts in relation to two samples and as good practice it is advisable to also consider the results of additional test. An additional test assessing the null hypothesis of the existence of a unit root is the PP test. The lag length and model specifications are in line with the ADF test approach. The PP test results are presented in Table 4. The pattern of results is confirming the ADF test results.

The KPSS test requires both lag length specification and model specification. As previously outlined, the choice to include a deterministic term is not evident from inspecting the data plots. The two model specifications are hence both evaluated. Kwiatkowski et al. (1992) suggest $\sqrt{T}$, where T is the sample size, is an appropriate lag length. In addition to the suggested lag length, results with lag length equal to one and ten are considered to determine whether the results are consistent across lag length choice. The results for all samples, model specifications, and lag length choices are presented in Table 5. The results for the EUA samples are only consistent across lag length and model specification for the full sample and sample three. For these samples, the KPSS test clearly rejects the null of stationarity. Sample 2 and Sample 4 of the EUA series show p-values of below 10% but above 5%, suggesting the null cannot be rejected at the 5% significance level, but it can be rejected at the 10% significance level. Lower leg choices imply clear rejections of the null, and the assumption of a trend for EUA Sample 4 does not appear suitable when considering the graphical evidence in Figure 10. Therefore, the results for sub-sample two and four are leaning towards the suggestion that the time series are non-stationary in line with the EUA Full Sample and Sample 3. P-values for EUA Sample 1 suggest results in favor of the stationarity assumption for the model specification that includes a deterministic trend and for both lag length 10 and 22 $(= \sqrt{T})$. This result is consistent with the ADF test findings, although the magnitude of the deterministic trend is questionable.

The results for the TTF samples are consistent across lag length and model specification for the Full Sample, Sample 3, and Sample 4, while Sample 1 and Sample 2 show inconsistencies. The former samples' results suggest to reject the null of stationarity in line with ADF test results. TTF Sample 1 and Sample 2 do not appear to include a trend term according to Figures 5 and 7. Based on that and the results of the ADF test, it appears reasonable to assume TTF Sample 2 is non-stationary. For TTF Sample 1, both the results, including and excluding a deterministic trend, cannot reject the null at the 5% significance level. However, the results are not consistent across lag length choice and the conclusion of stationarity would be inconsistent with results of the ADF test.

The KPSS test results for the log return series are not reported but all combinations of lag length and model specification for all samples of the EUA and TTF time series, except TTF sample one, are not significant at the 5% level, implying the first differences of the original time series are stationary according to the KPSS test. The TTF sample one results indicate p-values just below 5% for the model specification with a deterministic trend and lag length 10 and 22 ($= \sqrt{T}$).

In summary, the results of the unit root and stationarity tests of the price and log return series suggest that the price time series are integrated of order one, i.e., $I(1)$. This result is in line with expectations drawn from the graphical examination. Noteworthy is that the EUA Sample 1 series is potentially trend stationary while other inconsistencies across tests and model specifications did not rise to a level that would reasonably overturn the expectations from the graphical evidence. Therefore, it is appropriate to move forward with the cointegration analysis under the Johansen approach.

4.3 Cointegration Analysis

In this subsection, the results of the cointegration test under the Johansen approach are presented, and consequently, it is shown whether there is a long-term relationship between EUA and TTF prices. The results further indicate whether the relationship has changed over time. The cointegration test under the Johansen approach requires lag length and model specification for accurate results. The lag length is determined via information criteria for the VAR model in first differences, i.e., the log returns. A VAR(p) model is fitted for $p = \{1; 2; ..; 10\}$. For each model, the AIC and BIC are determined. Both the AIC and BIC suggest $p = 1$ has the best model fit. Therefore, the VEC model should have a lag length of $q = p - 1 = 0$. Johansen (1995) and Juselius (2006) suggest a lag length of $q = 2$ is sufficient for most applications. The AIC and BIC specified lag length is below the suggestion so that one can assume the number of parameters is not the issue. However, one may presume the autocorrelation is not appropriately captured with a small lag length specification. The innovation time series of the VAR(p) model are checked for autocorrelation to confirm the appropriateness of the lag length choice. The autocorrelation functions of the innovations show that there is little residual autocorrelation in the innovation series, and hence the lag length choice is appropriate.

In the next step, the model specification for the Johansen test is to be determined. As there is no theory on the nature of the relationship between the two variables and inconclusive evidence on the deterministic terms in the levels of the data, three out of the five models Johansen (1995) suggests, are considered. The three cointegration terms are described by Equations 18,19, and 20 and in the following referred to H1*, H1, and H*, respectively. The resulting VEC(q) models according to Equation 11 with $q = 0$ reduce to Equations 26, 27, and 28, respectively.

$$\Delta y_t = A(B' y_{t-1} + c_0) + \varepsilon_t \tag{26}$$

$$\Delta y_t = A(B' y_{t-1} + c_0) + c_1 + \varepsilon_t \tag{27}$$

$$\Delta y_t = A(B' y_{t-1} + c_0 + d_c t) + c_1 + \varepsilon_t \tag{28}$$

The results of the Johansen test for the three specifications are presented in Table 6. The results for both the trace test and maximum eigenvalue test are presented. The tests produce the same decision outcome across samples and models. The results show no statistically significant long-term relationship between emission allowance and natural gas in the Full Sample as well as Samples 1 to 3. The results show that C in Equation 13 has a rank of zero, implying there is no cointegration relationship, and the VAR(1) model in first differences is the appropriate multivariate model of the respective time series. In contrast, the results for Sample 4, i.e., the calendar year 2019, show that there is a significant cointegration relationship using models H1* and H1. The results are significant at the 10% and 5% level, respectively. A likelihood ratio test is employed to check whether one of the models is more appropriate than the others. The results of the pairwise likelihood ratio test for Sample 4 are presented in Table 7. The results suggest that the most restrictive model H1* in line with Equation 26 can not be rejected in favor of the less restricted H1 and H*. The least restrictive model H* is clearly not superior to the less restrictive models. This result also provides additional comfort the cointegration relationship is statistically significant as the H* does not suggest a relationship exists compared to the other two models that appear to be more appropriate according to the likelihood ratio test.

The non-significant cointegration relationship in the full and earlier sample periods versus the significant relationship in 2019 suggests that the relationship between the variables has changed over time. This result is supported by the hypothesis that the sustained high price levels of emission allowances are robust enough to create a link to natural gas prices as the relative attractiveness of fuel switching increases with increased emission price levels. In addition, regulatory changes in the EU ETS and in particular the introduction of the Market Stability Reserve (MRS) (European Union, 2015) in 2019 make the emission allowances market less susceptible to structural oversupply that had left the emission market at depressed prices throughout the second and large parts of the third trading period from 2008 to 2018. The relatively higher attractiveness of fuel switching and hence the demand for natural gas is not only assumed to be driven by emission prices by making it cheaper to use natural gas compared to coal for power generation but also by direct regulations aimed at reducing or fully abolishing coal as a fuel source in the long term. The latter is underpinned by coal phase-out plans on national levels within the EU. Such laws have been adopted in recent times as environmental awareness has gained mainstream traction and governments, supranational institutions, and private corporations have set ambitious goals to reduce emissions. Of course, such ambitious goals are also driving innovation that may make fossil fuel-based power generation completely obsolete in the future. It is further noteworthy that the global oil and gas market has also experienced a significant shift in the past decade. The exploration of shale gas in the US has transformed the local energy mix away from coal to the abundant and hence cheap natural gas as a fuel source. This shift consequently impacted energy markets globally with an increased supply of no longer needed coal while natural gas remained on the local market due to the lacking technology to transport large volumes. In the absence of meaningful innovation or regulatory breaks, one may assume the cointegration relationship between EUAs and TTFs, as suggested by the results for 2019 will continue into the future.

The dynamics of the cointegration relation are described by the VEC model that is estimated via maximum likelihood methodology in the Johansen approach. The estimated model for Sample 4 and the H1* model specification is described by

$$\Delta y_t = \begin{bmatrix} -0.1281 \\ -0.1292 \end{bmatrix} \left(\begin{bmatrix} 0.4755 & 0.2921 \end{bmatrix} y_{t-1} + (-15.56) \right) + \varepsilon_t \tag{29}$$

in line with Equation 26. Within the Johansen approach, the vectors A and B are not separately estimated but rather the cointegration matrix C as described by Equation 12 is estimated. The estimation results and standard errors, t-stats, and p-values are presented in Table 8. The results show that the estimates are highly significant. In an additional step, the adjustment speeds of vector A as shown in Equation 29 as $\begin{bmatrix} -0.1281 \\ -0.1292 \end{bmatrix}$ are tested in the Johansen framework to determine whether the coefficients are significantly different from zero. The test set up is comparing a restricted with an unrestricted model, and the test statistics are formed by Equation 16. The restriction is imposed on each element of A separately such that the restricted Matrix $A_r^1 = \begin{bmatrix} 0 \\ -0.1292 \end{bmatrix}$ and $A_r^2 = \begin{bmatrix} -0.1281 \\ 0 \end{bmatrix}$ are used in the estimation of the log-likelihood of the restricted model. The null hypothesis states the restriction holds, i.e., the restricted model is true versus the unrestricted model of the alternative hypothesis. If one of the elements of A was determined to be zero, then the respective variable is not adjusting from disequilibrium in the cointegration relationship. In the context of two variables, that means if one was zero, the other variable would be 100% responsible for the system to return to equilibrium. The results of the adjustment speeds test in the Johansen framework are shown in Table 9 and show that both the adjustment speed parameters of the EUA and the TTF series are not zero as both restricted models are clearly rejected with p-values below 1%. Therefore, both variables contribute to the adjustment to disequilibrium. In the next step, it is determined whether the EUA or TTF time series is leading the price discovery, meaning which market incorporates information first.

4.4 Price Discovery

Price discovery measures help understand the directionality of the relationship between variables. An initial hypothesis is that the EUA market is incorporating information first as the EUA price influences the demand for TTF by virtue of making fuel switching more attractive and therefore implying higher demand for natural gas. On the other hand, one may presume that TTF prices are directly making it more or less attractive for power producers to switch fuels and that the TTF price determines the fuel switching decision that will consequently influence the power producer's requirement to purchase EUAs due to lower emissions output (in the case of a positive fuel switching decision from coal to gas) and hence impact the demand of EUAs. A first measure to check the price discovery is the component share as described in Equation 23, and results are presented in Table 10. As the two adjustment speeds are relatively close in magnitude, the resulting measures are rather large. However, it is clear that the component share of the EUA series is larger in absolute values than the component share of the TTF series. This result implies that EUAs are leading the price discovery process according to the component share measure. The second measure, the information share, is described by Equations 24 and 25, and results are presented in Table 10. The average of the lower and upper IS measures indicate that the EUA series is leading the price discovery given the average IS share is above 50%. However, it is worth highlighting that the results are very close to 50%, and the lower and upper bounds cross over the 50% mark for both variables. As Baillie et al. (2002) suggest, the spread between the upper and lower IS measure are wide when there is a large and significant correlation between the residuals of the VECM. Checking the residual series indeed shows that they are significantly correlated with correlation equal to $p = 0.29$. Although both the CS and IS measures are consistent and indicate the EUA series is leading the price discovery process, it is equally clear that the results are close, and one may conclude there is insufficient evidence to support the hypothesis that EUA prices incorporating information first.

4.5 Comparison with other papers

The existing literature shows varying cointegration results for the pilot phase of the EU ETS while showing more consistency for the second trading period. While Bunn and Fezzi (2007) find a cointegration relationship among emission allowances, coal, gas, and electricity prices, Rickels et al. (2007) does not find evidence to suggest there is a cointegration relationship among emission allowances and energy inputs. While using the data from the same time period the authors of the two papers use different proxies for both coal and natural gas. Extending the time period and incorporating coal and natural gas fundamentals via dark and spark spreads, respectively, Bredin and Muckley

(2011) show that the cointegration relationship among fundamental drivers of emission allowances is stronger in the second trading phase of the EU ETS than the first. Creti et al. (2012) is supporting this notion by showing that switching costs do not have a long-run relationship with emission allowances in the first but in the second trading period. Compared to the existing literature, this thesis is considering a different natural gas proxy and time period with data spanning the third trading period of the EU ETS and hence incorporating the latest market and regulatory changes. However, the result of this thesis that there is a cointegration relationship in 2019 potentially indicates that in market environments that would suggest fundamentals are the driving forces, the cointegration relationship between emissions and respective energy prices appears to be present.

5 Conclusion

This thesis examines the long-run relationship between emission allowances and natural gas. The Johansen cointegration testing framework is employed to establish whether the time series are cointegrated when considering three model specifications and five different considered sample periods. Preliminary testing of the order of integration of the respective univariate time series is conducted in order to verify the prerequisites for cointegration analysis are indeed met. The results show the time series are integrated of order one, i.e., $I(1)$ across time periods, and hence prerequisites for the Johansen test are met. In order to accommodate potential deterministic terms and different cointegration relationships, three model specifications of the error correction terms are analyzed. The cointegration test is conducted both via trace test and maximum eigenvalue test within the Johansen framework. The cointegration tests show that there is no evidence of a long-run relationship for the Full Sample period and Sample periods 1 to 3 independent of the model specification. In contrast, Sample 4, spanning the time period from January 2019 to December 2019, shows there is a cointegration relationship when considering two out of three model specifications. The results are consistent across trace and maximum eigenvalue tests. A likelihood ratio test confirms that the most restrictive model specification is the most appropriate choice. This specification shows that the cointegration relationship is significant at the 10% significance level. In an effort to understand the price discovery mechanics, the component share and information share measures are calculated and analyzed. The results suggest that EUAs are leading the price discovery while noting the close results and correlation in the innovation process, weakening that conclusion.

Considering the results suggest that only in the most recent year there has been a statistically significant long-run relationship between emission allowances and natural gas, it is a logical conclusion that the underlying relationship between the two variables has changed over time. However, the driver of this change is not explicitly identified as part of the scope of this thesis. However, it is noteworthy that the EUA market has suffered a long period of structural oversupply and, consequently, depressed prices that suggest that the EUA market was removed from fundamental driving forces. At the beginning of 2019, the Market Stability Reverse (MSR) was introduced to address precisely this issue of oversupply and increase the functionality of the EUA market. The 2019 price levels in the EUA market are a positive signal the MSR is working and expected to work going forward. Considering the high price levels in 2019 with the finding that there is a long-run relationship between emission allowances and natural gas, one may further suspect that the EU ETS is effective in incentivizing market participants to switch fuel from higher carbon-intensive coal to gas and therefore reduce emitted emissions. The results from the price discovery analysis support this hypothesis but do not conclusively establish this result.

It is clear from both the results of this thesis and the existing literature regarding earlier trading periods that a continued assessment of the latest market data is required to confirm the conclusions of this thesis. The recent market turmoil due to the Coronavirus and the oil price dislocation offer additional opportunities to test the robustness of the cointegration relationship among emission allowances, economic activity, and energy market fundamentals, including natural gas prices.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*(6), 716–723.

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

Baillie, R. T., G. G. Booth, Y. Tse, and T. Zabotina (2002). Price discovery and common factor models. *Journal of Financial Markets 5*, 309–321.

Bredin, D. and C. Muckley (2011). An emerging equilibrium in the EU emissions trading scheme. *Energy Economics 33*, 353–362.

Brooks, C. (2002). *Introductory econometrics for finance*. Cambridge University Press.

Bunn, D. W. and C. Fezzi (2007). Interaction of European carbon trading and energy prices. Technical report, Nota di Lavoro.

Christiansen, A. C., A. Arvanitakis, K. Tangen, and H. Hasselknippe (2005). Price determinants in the EU emissions trading scheme. *Climate Policy 5*(1), 15–30.

Creti, A., P.-A. Jouvet, and V. Mignon (2012). Carbon price drivers: Phase I versus Phase II equilibrium? *Energy Economics 34*(1), 327–334.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*, 427–431.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: Journal of the Econometric Society 55*(2), 251–276.

Engle, R. F. and B. S. Yoo (1987). Forecasting and testing in co-integrated systems. *Journal of Econometrics 35*, 143–159.

European Union (2003). Directive 2003/87/EC of the European Parliament and of the Council of 13 October 2003 establishing a system for greenhouse gas emission allowance trading within the Union and amending Council Directive 96/61/EC. *Official Journal of the European Union L275*, 32–46.

European Union (2015). Decision (EU) 2015/1814 of the European Parliament and of the Council of 6 October 2015 concerning the establishment and operation of a market stability reserve for the Union greenhouse gas emission trading scheme and amending Directive 2003/87/EC. *Official Journal of the European Union L 264*, 1–5.

Gonzalo, J. and C. Granger (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics 13*(1), 27–35.

Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *The Journal of Finance 50*(4), 1175–1199.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control 12*(2-3), 231–254.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press.

Johansen, S. and K. Juselius (1990). Maximum likelihood estimation and inference on cointegrationâwith applications to the demand for money. *Oxford Bulletin of Economics and Statistics 52*, 169–210.

Juselius, K. (2006). *The cointegrated VAR model: Methodology and applications*. Oxford University Press.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, Y. Shin, et al. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics 54*, 159–178.

Mansanet-Bataller, M., A. Pardo, and E. Valor (2007). CO_2 prices, energy and weather. *The Energy Journal 28*, 73–92.

Phillips, P. C. B. and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika 75*(2), 335–346.

Rickels, W., V. Duscha, A. Keller, S. Peterson, et al. (2007). The determinants of allowance prices in the European emissions trading scheme: Can we expect an efficient allowance market 2008? Technical report, Kiel Working Paper.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*, 461–464.

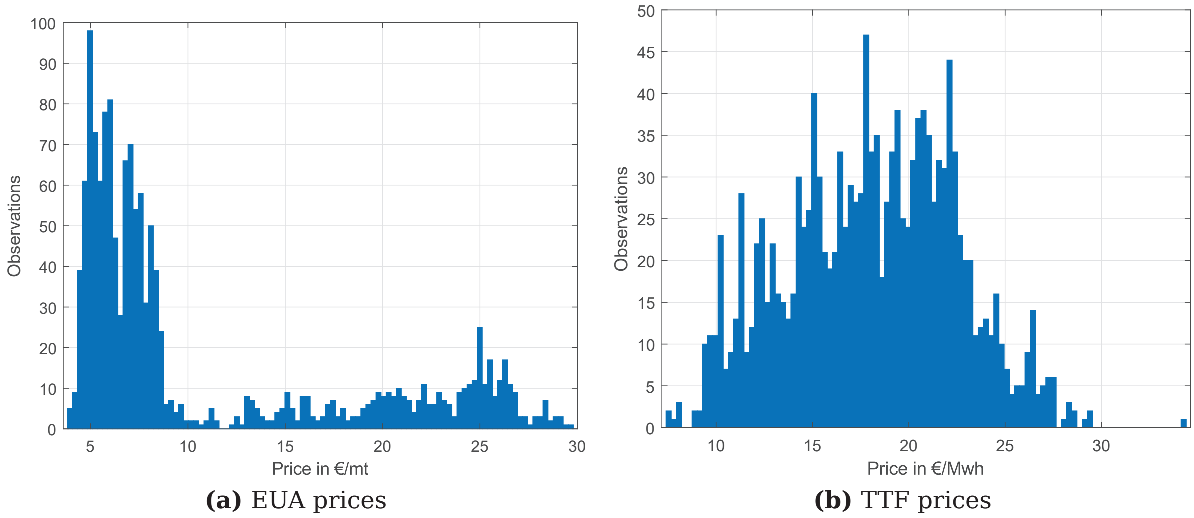


Figure 1: Histograms of the daily EUA prices (panel a) and daily TTF prices (panel b) from January 01, 2014 to December 31, 2019

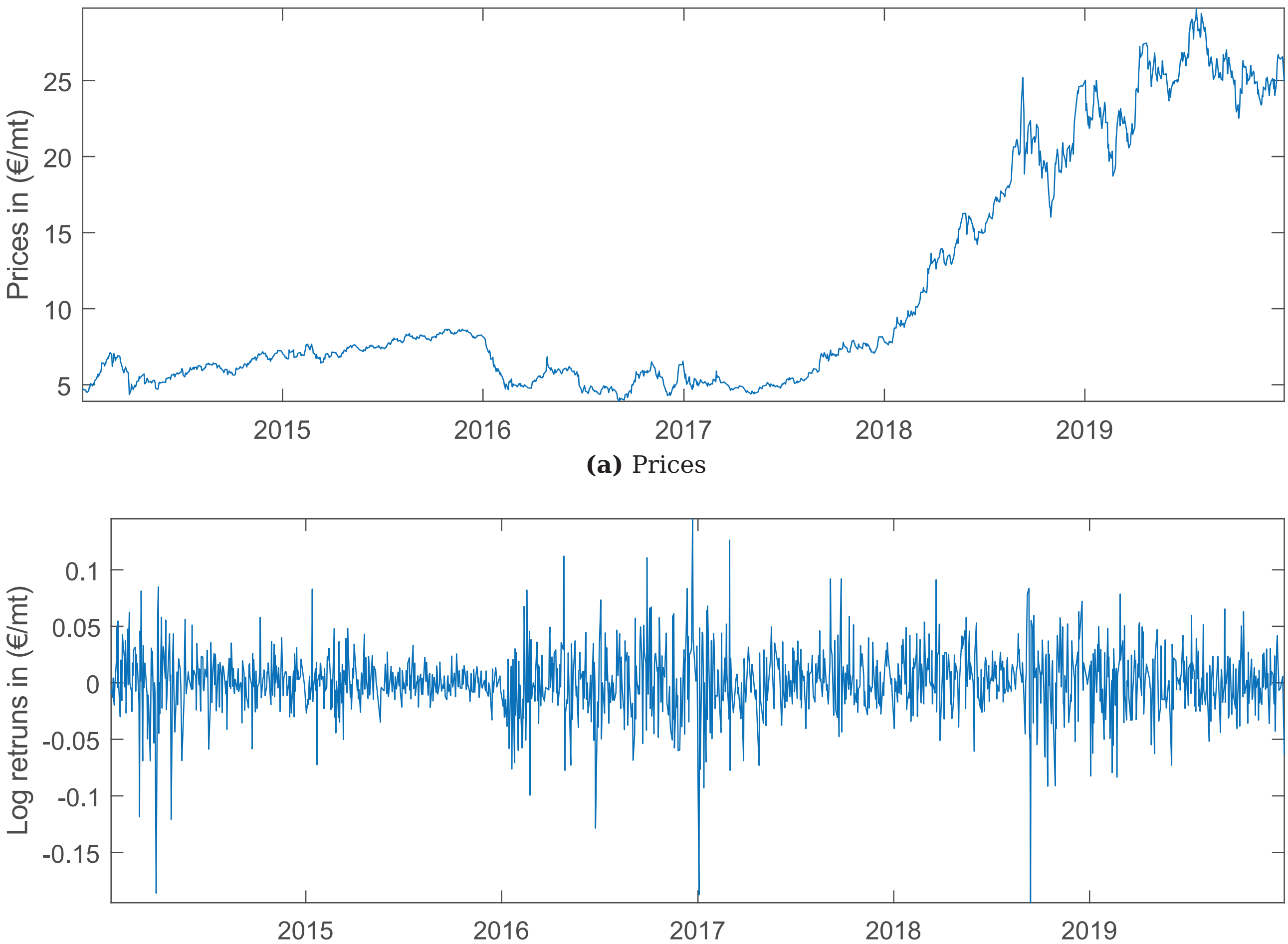
Table 1: Descriptive statistics for the daily EUA (panel a) and TTF (panel b) price series for the Full Sample from January 01, 2014 to December 31, 2019, Sample 1 from January 01, 2014 to December 31, 2016, Sample 2 from January 01, 2016 to December 31, 2017, Sample 3 from January 01, 2018 to December 31, 2019, and the Sample 4 from January 01, 2019 to December 31, 2019.

(a) EUA prices

Time Period	Full sample	Sample 1	Sample 2	Sample 3	Sample 4
Mean	10.97	6.83	5.62	20.36	24.81
Minimum	3.91	4.35	3.91	7.62	18.72
Maximum	29.76	8.65	8.17	29.76	29.76
Standard deviation	7.50	1.07	1.00	5.72	2.19

(b) TTF prices

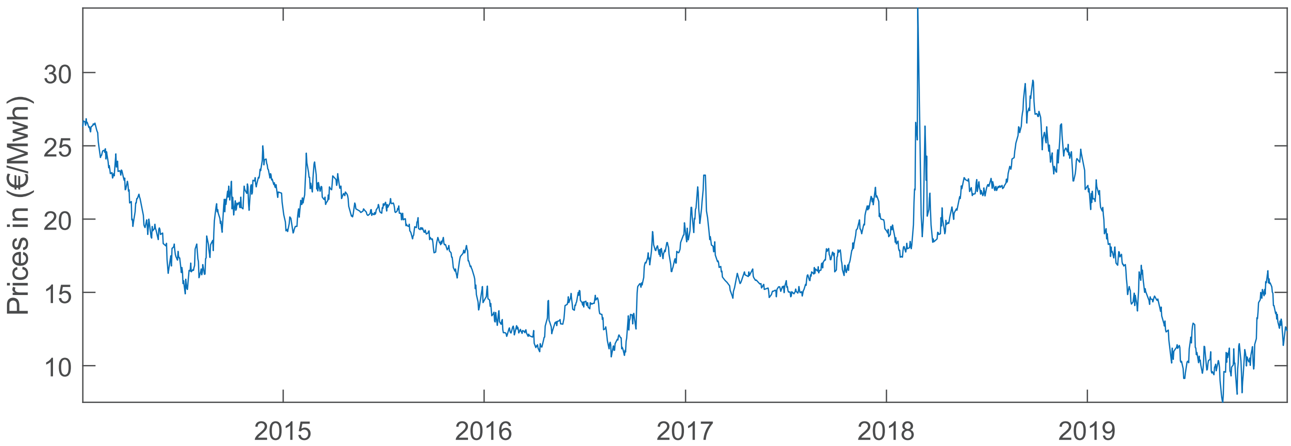
Time Period	Full sample	Sample 1	Sample 2	Sample 3	Sample 4
Mean	18.08	20.43	15.63	18.09	13.60
Minimum	7.50	13.80	10.60	7.50	7.50
Maximum	34.40	26.85	23.00	34.40	22.68
Standard deviation	4.39	2.58	2.76	5.65	3.68



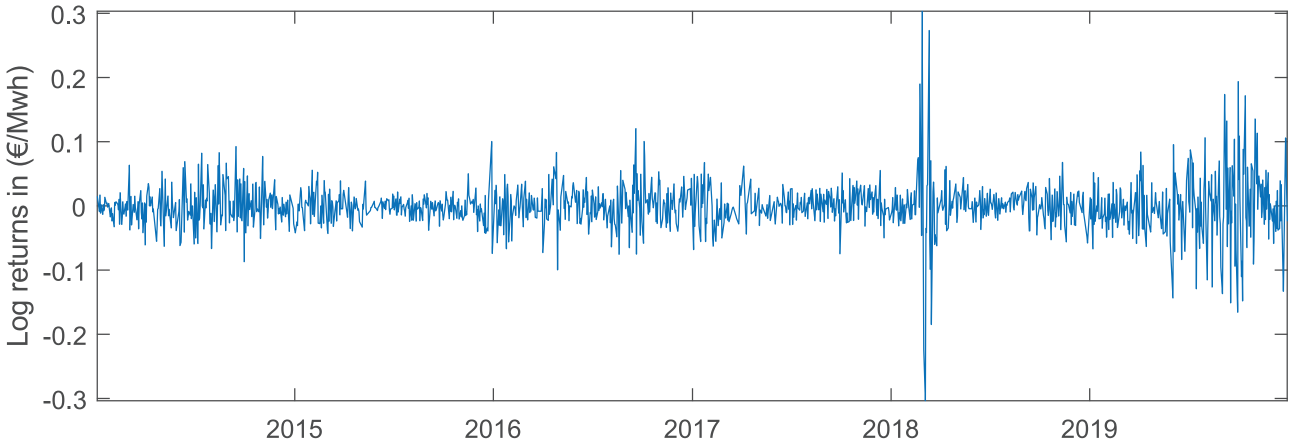
(a) Prices

(b) Log returns

Figure 2: Daily EUA spot prices and EUA log returns from January 01, 2014 to December 31, 2019

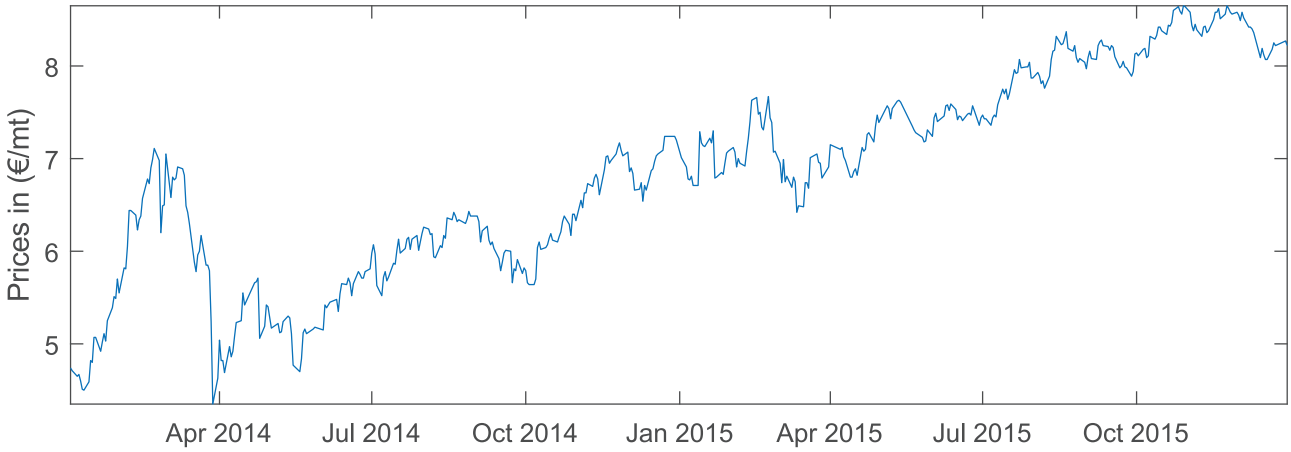


(a) Prices

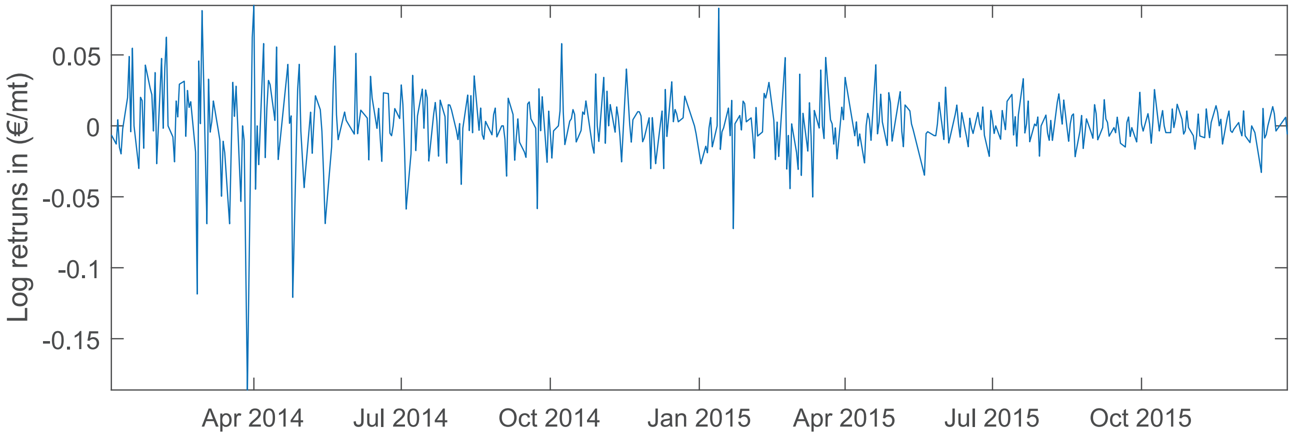


(b) Log returns

Figure 3: Daily TTF futures prices and log returns from January 01, 2014 to December 31, 2019

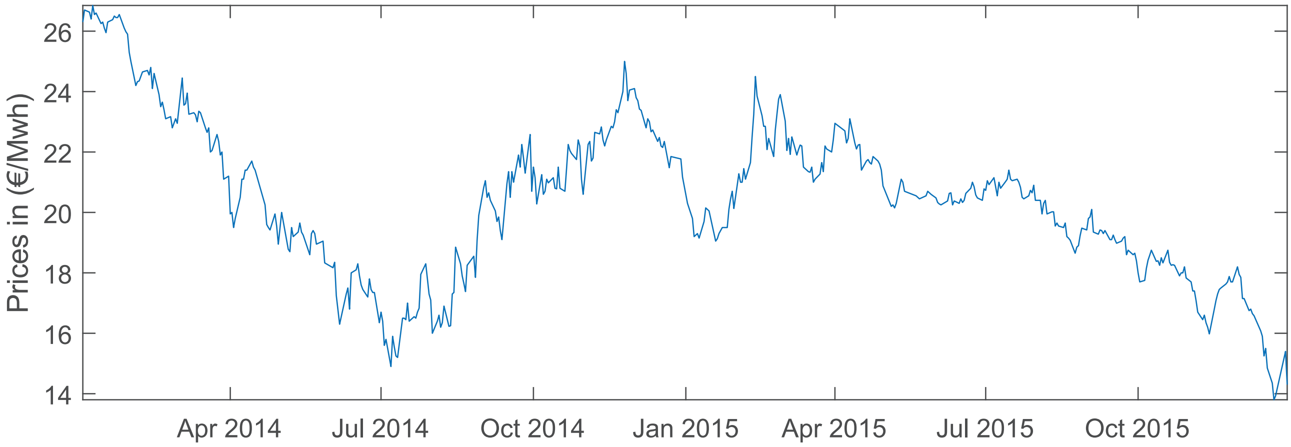


(a) Prices

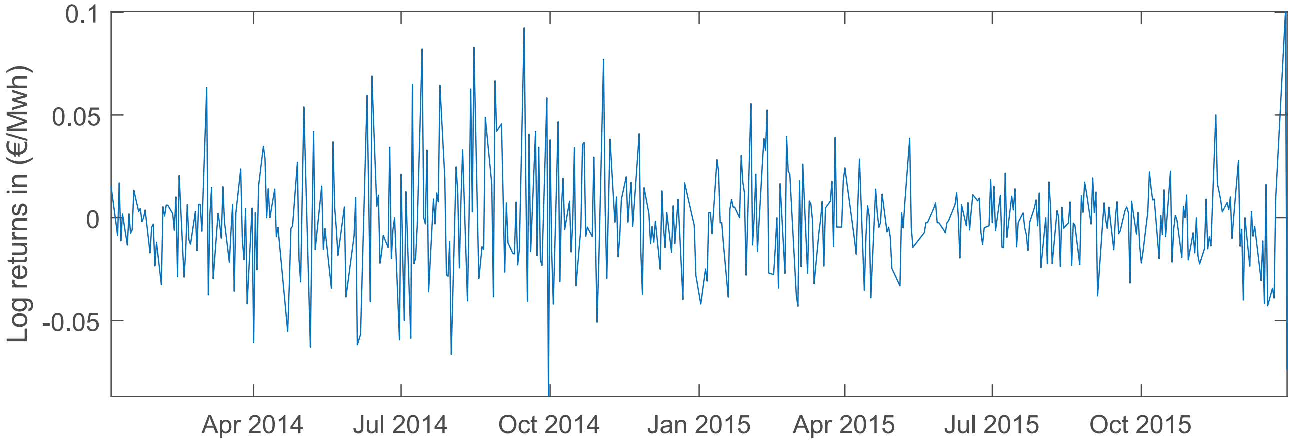


(b) Log returns

Figure 4: Daily EUA spot prices and EUA log returns from January 01, 2014 to December 31, 2015

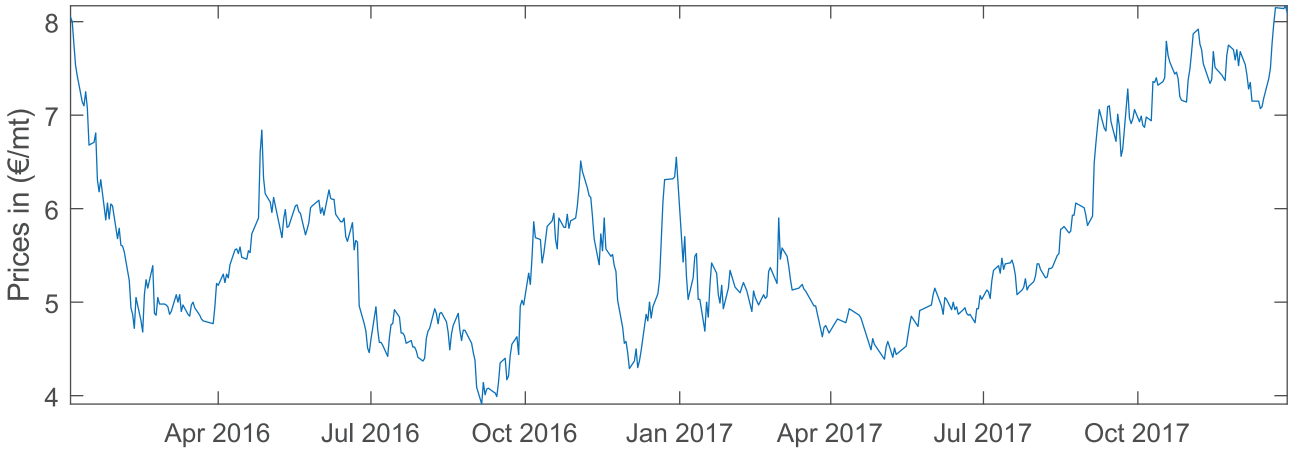


(a) Prices

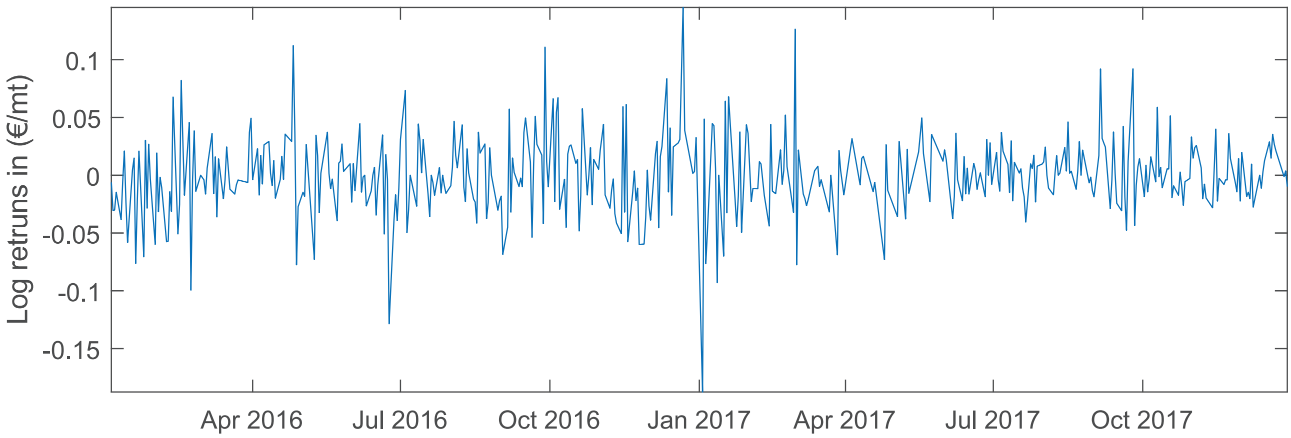


(b) Log returns

Figure 5: Daily TTF futures prices and log returns from January 01, 2014 to December 31, 2015

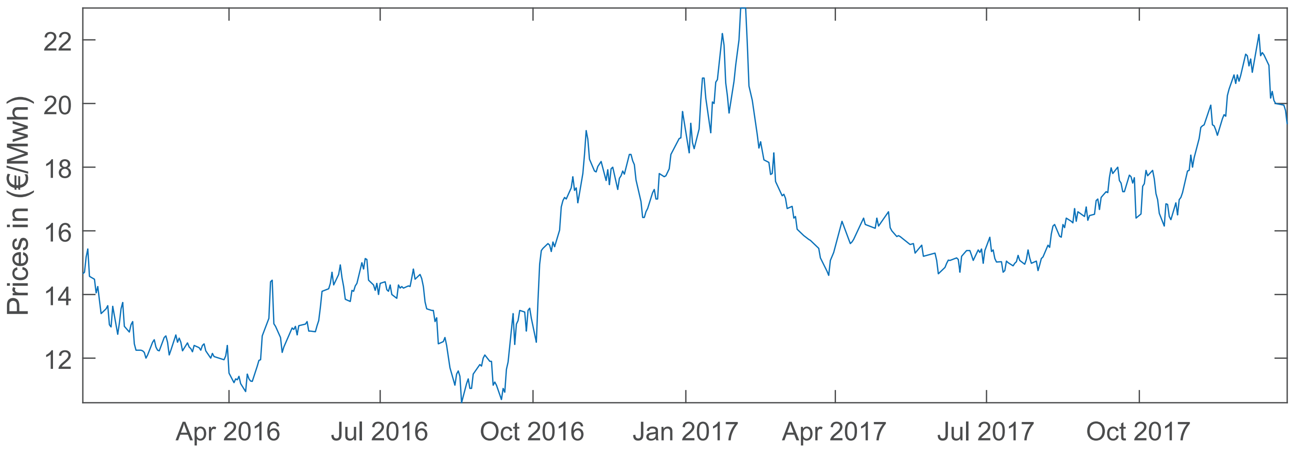


(a) Prices

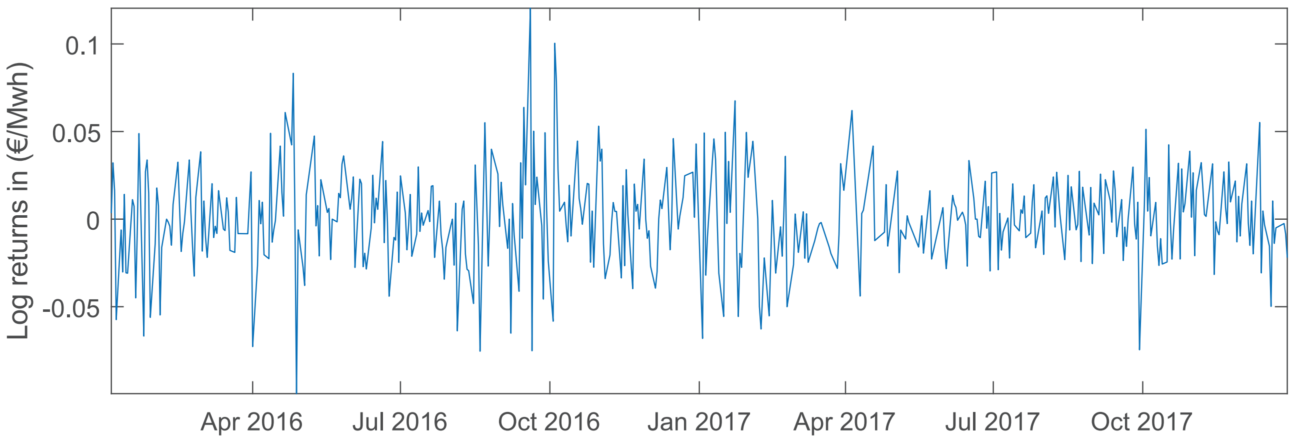


(b) Log returns

Figure 6: Daily EUA spot prices and EUA log returns from January 01, 2016 to December 31, 2017

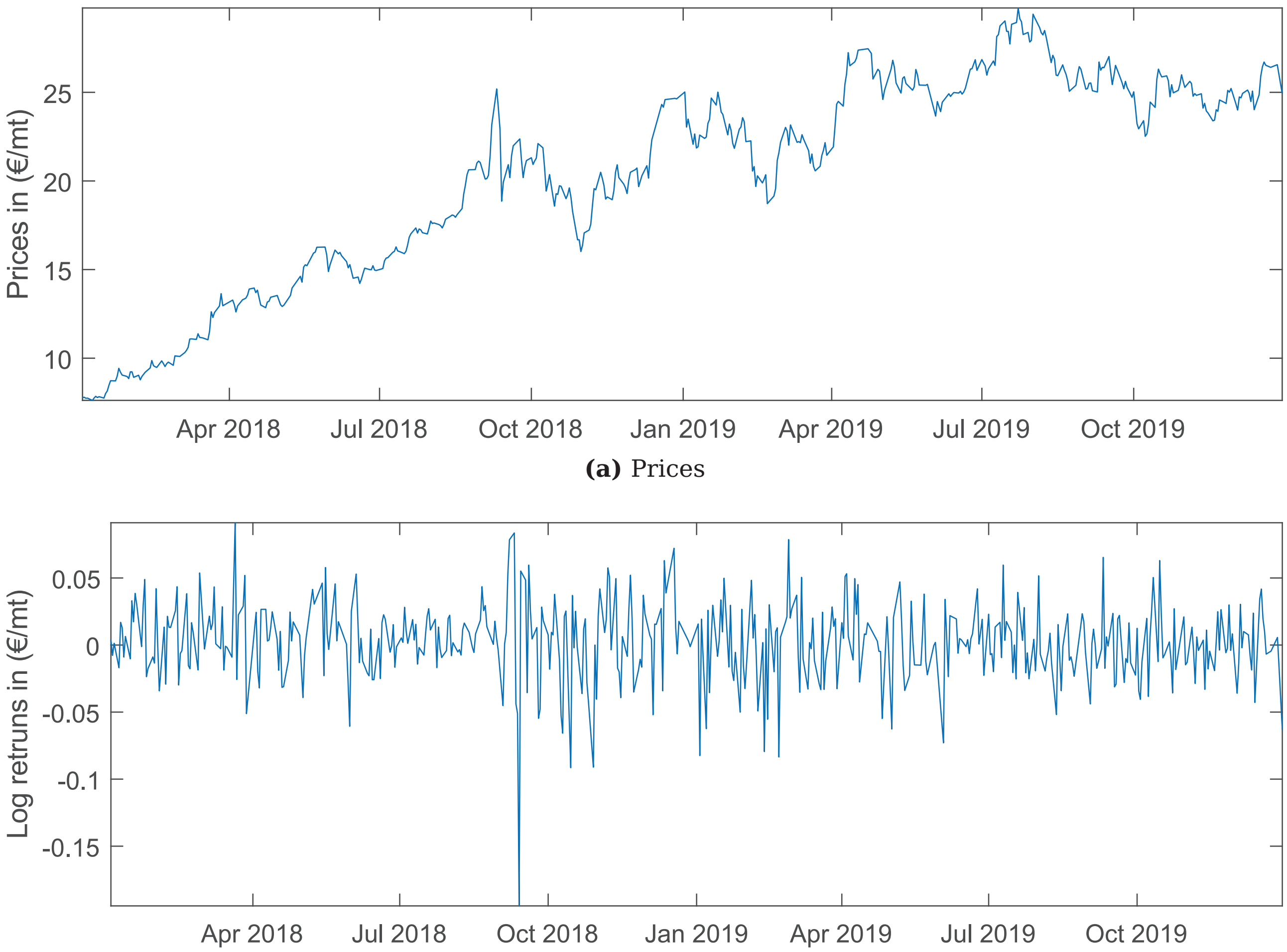


(a) Prices



(b) Log returns

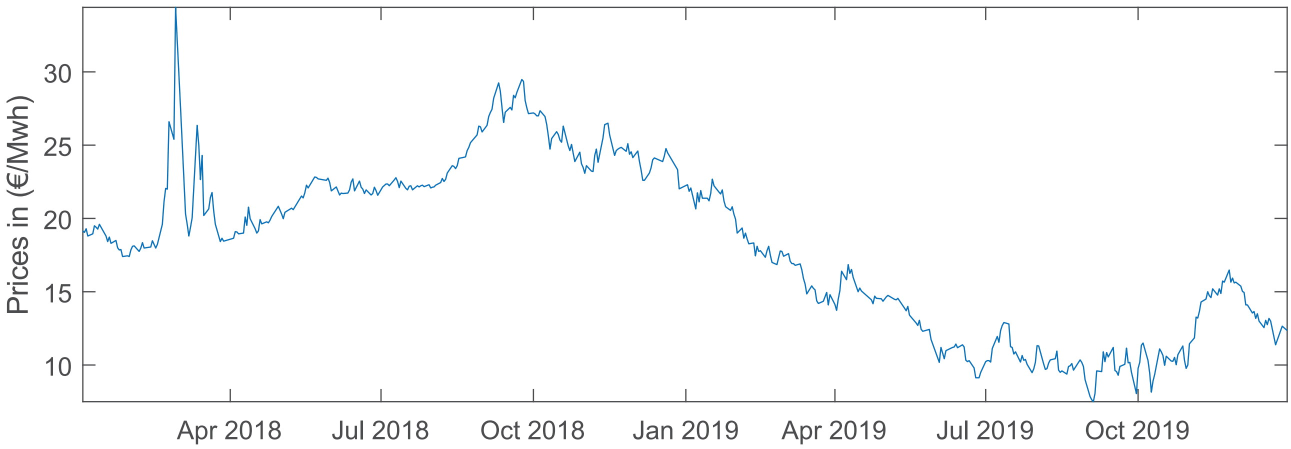
Figure 7: Daily TTF futures prices and log returns from January 01, 2016 to December 31, 2017



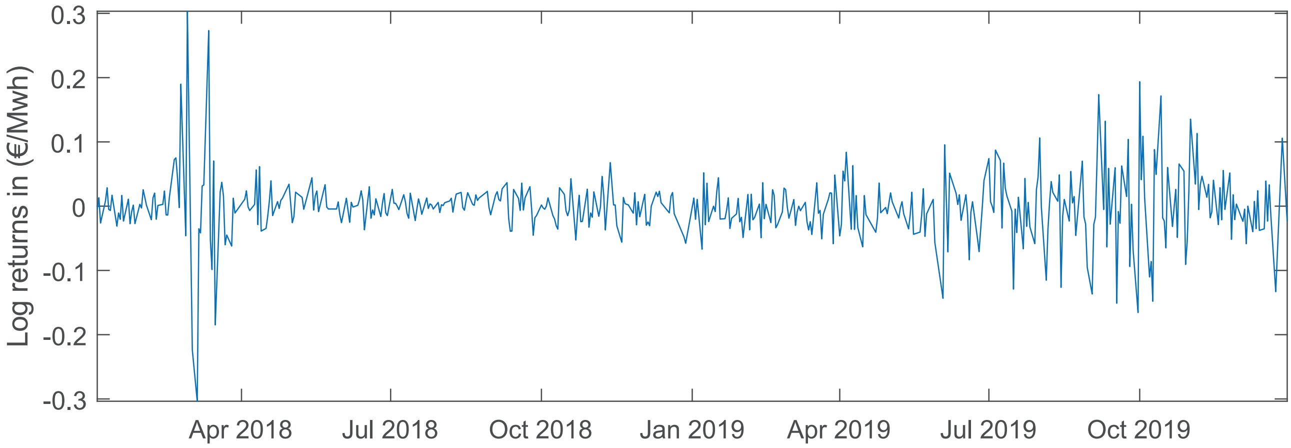
(a) Prices

(b) Log returns

Figure 8: Daily EUA spot prices and EUA log returns from January 01, 2018 to December 31, 2020

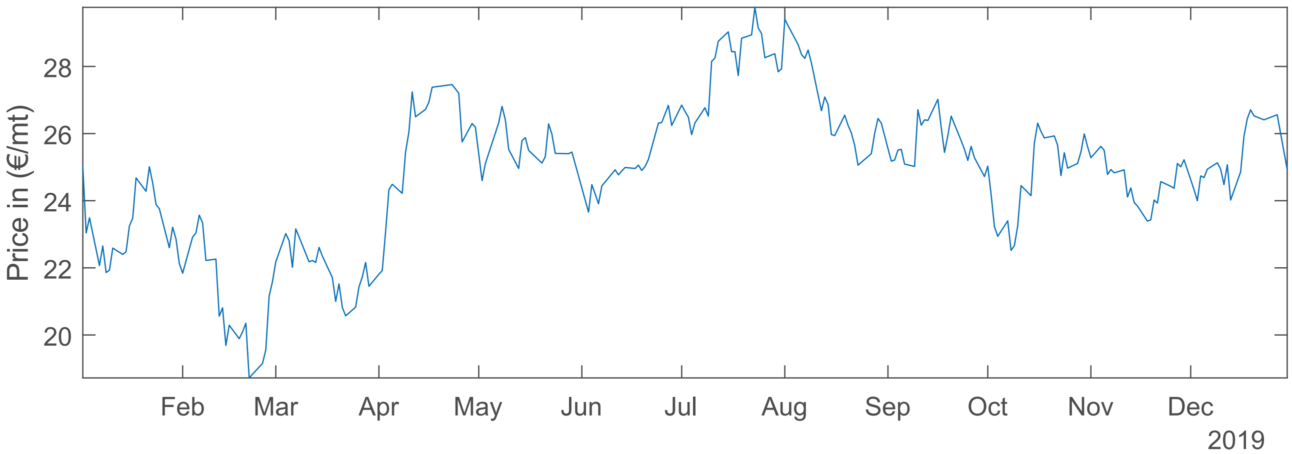


(a) Prices

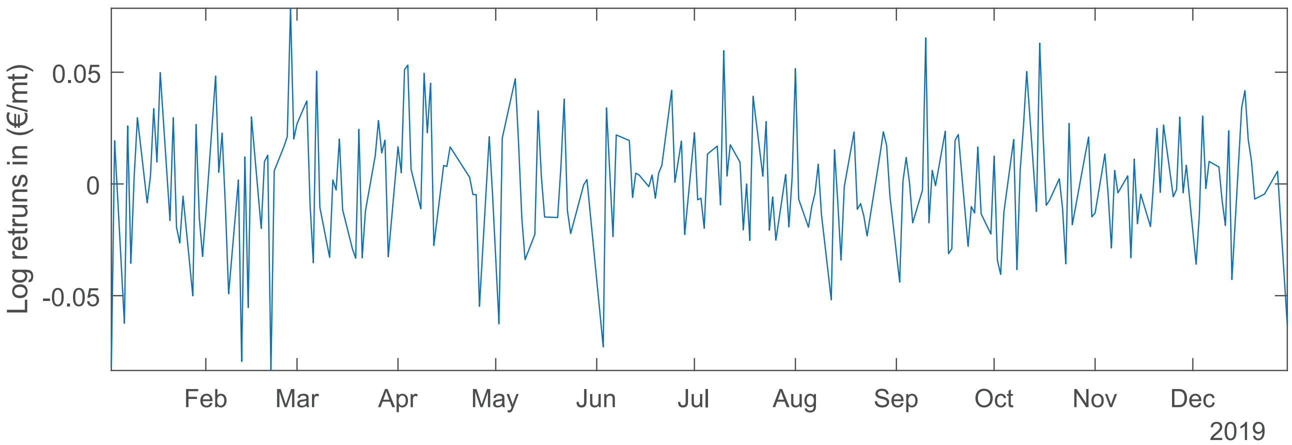


(b) Log returns

Figure 9: Daily TTF futures prices and log returns from January 01, 2018 to December 31, 2020

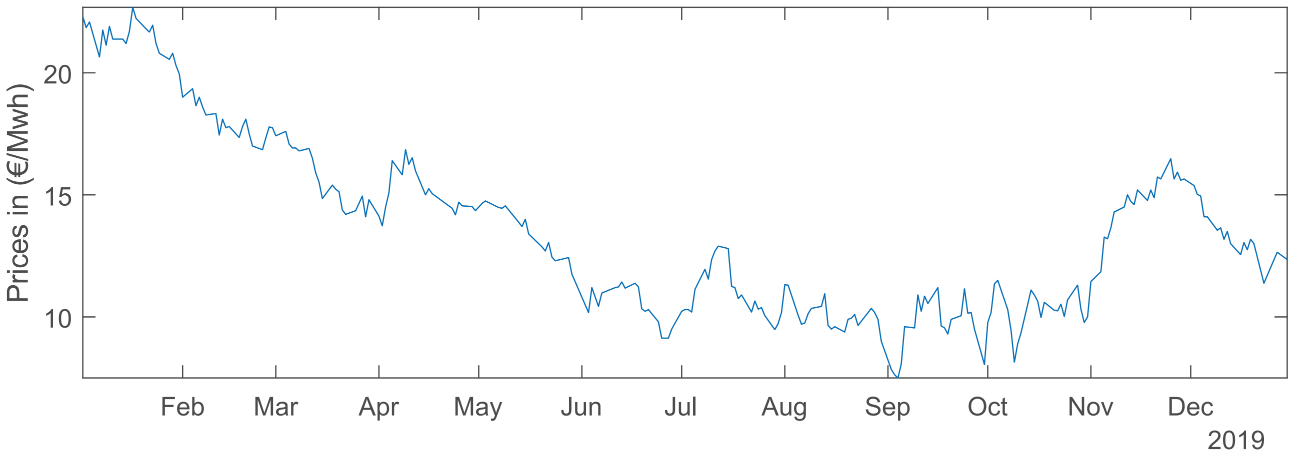


(a) Prices

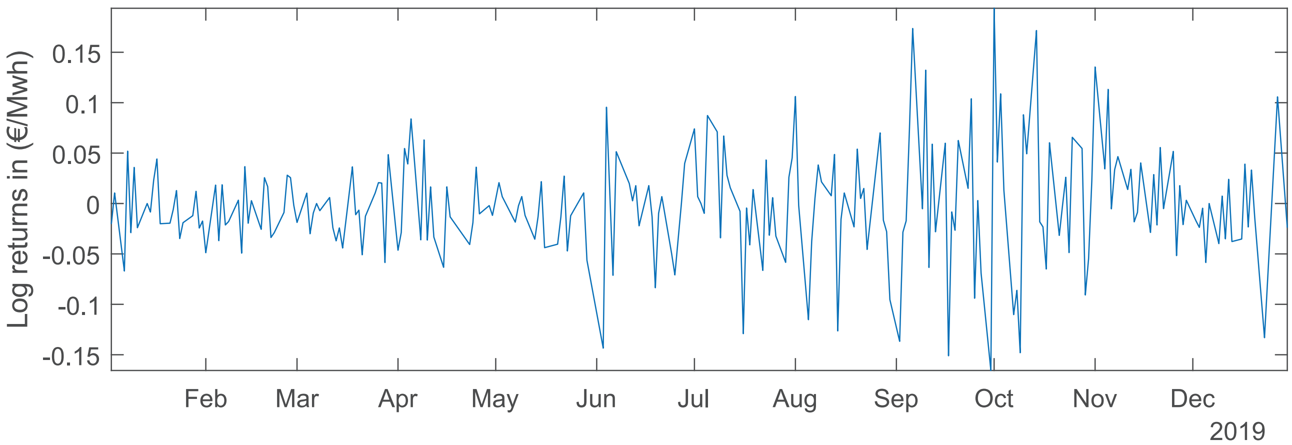


(b) Log returns

Figure 10: Daily EUA spot prices and EUA log returns from January 01, 2019 to December 31, 2020



(a) Prices



(b) Log returns

Figure 11: Daily TTF futures prices and log returns from January 01, 2019 to December 31, 2020

Table 2: Augmented Dickey-Fuller test results, as measured by p-values, for the daily EUA (panel a) and TTF (panel b) price series for the Full Sample and Samples 1 to 4 as previously specified. The model specification AR refers to Equation 4 where $c = \delta = 0$. The model specification ARD refers to Equation 4 where $\delta = 0$. The model specification TS refers to Equation 4. Results for the t-test are presented for all model specifications and results for the F-test are only presented for the ARD and TS specification. For the ARD test the F-test assesses the joint hypothesis that $\phi - 1 = 0$ and $c = 0$ and for the TS specification the F-test assesses the joint hypothesis that $\phi - 1 = 0$ and $\delta = 0$. The lag length specifications are based on the minimum respective information criteria.

(a) EUA prices

			t-test			F-test	
		Lags	AR	ARD	TS	ARD	TS
Full sample	AIC	9	0.96	0.96	0.87	0.69	0.86
	BIC	0	0.92	0.93	0.79	0.78	0.82
Sample 1	AIC	0	0.88	0.36	0.02	0.39	0.03
	BIC	0	0.88	0.36	0.02	0.39	0.03
Sample 2	AIC	0	0.53	0.29	0.09	0.47	0.02
	BIC	0	0.53	0.29	0.09	0.47	0.02
Sample 3	AIC	7	0.92	0.23	0.48	0.11	0.43
	BIC	0	0.86	0.26	0.24	0.20	0.23
Sample 4	AIC	6	0.71	0.20	0.47	0.33	0.60
	BIC	1	0.66	0.22	0.44	0.36	0.59

(b) TTF prices

			t-test			F-test	
		Lags	AR	ARD	TS	ARD	TS
Full sample	AIC	10	0.19	0.17	0.42	0.19	0.57
	BIC	7	0.18	0.19	0.46	0.22	0.61
Sample 1	AIC	10	0.13	0.39	0.62	0.36	0.78
	BIC	1	0.11	0.37	0.59	0.30	0.75
Sample 2	AIC	9	0.74	0.38	0.26	0.55	0.38
	BIC	0	0.75	0.57	0.37	0.76	0.48
Sample 3	AIC	10	0.38	0.78	0.60	0.94	0.71
	BIC	7	0.37	0.76	0.59	0.93	0.71
Sample 4	AIC	4	0.07	0.17	0.83	0.13	0.57
	BIC	4	0.07	0.17	0.83	0.13	0.57

Table 3: Deterministic coefficient estimates of the ARD and TS model specifications under the Augmented Dickey-Fuller test for the daily EUA (panel a) and TTF (panel b) price series for the Full Sample and Samples 1 to 4 as previously specified. The lag length specifications are based on the minimum respective information criteria.

(a) EUA prices

		Lags	ARD constant	TS constant	TS trend
Full sample	AIC	9	0.02	0.00	0.0001
	BIC	0	0.02	0.00	0.0001
Sample 1	AIC	0	0.09	0.30	0.0004
	BIC	0	0.09	0.30	0.0004
Sample 2	AIC	0	0.10	0.12	0.0002
	BIC	0	0.10	0.12	0.0002
Sample 3	AIC	7	0.27	0.44	0.0007
	BIC	0	0.24	0.44	0.0010
Sample 4	AIC	6	1.12	1.23	0.0004
	BIC	1	1.03	1.16	0.0005

(b) TTF prices

		Lags	ARD constant	TS constant	TS trend
Full sample	AIC	9	0.15	0.17	0.0000
	BIC	0	0.15	0.17	0.0000
Sample 1	AIC	0	0.30	0.39	-0.0001
	BIC	0	0.29	0.38	-0.0001
Sample 2	AIC	0	0.21	0.34	0.0004
	BIC	0	0.16	0.29	0.0004
Sample 3	AIC	7	0.10	0.57	-0.0008
	BIC	0	0.11	0.57	-0.0008
Sample 4	AIC	6	0.29	0.20	0.0003
	BIC	1	0.29	0.20	0.0003

Table 4: Phillips-Perron test results, as measured by p-values, for the daily EUA (panel a) and TTF (panel b) price series for the Full Sample and Samples 1 to 4 as previously specified. The model specification AR refers to Equation 5 where $c = \delta = 0$. The model specification ARD refers to Equation 5 where $\delta = 0$. The model specification TS refers to Equation 5.The lag length specifications are based on the minimum respective information criteria.

(a) EUA prices

		Lags	AR	ARD	TS
Full sample	AIC	9	0.95	0.95	0.84
	BIC	0	0.92	0.93	0.79
Sample 1	AIC	0	0.88	0.36	0.02
	BIC	0	0.88	0.36	0.02
Sample 2	AIC	0	0.53	0.29	0.09
	BIC	0	0.53	0.29	0.09
Sample 3	AIC	7	0.88	0.27	0.31
	BIC	0	0.86	0.26	0.24
Sample 4	AIC	6	0.58	0.15	0.23
	BIC	1	0.58	0.18	0.31

(b) TTF prices

		Lags	AR	ARD	TS
Full sample	AIC	10	0.19	0.12	0.33
	BIC	7	0.19	0.12	0.32
Sample 1	AIC	10	0.11	0.42	0.62
	BIC	1	0.14	0.37	0.57
Sample 2	AIC	9	0.76	0.58	0.36
	BIC	0	0.75	0.57	0.37
Sample 3	AIC	10	0.37	0.66	0.44
	BIC	7	0.37	0.65	0.43
Sample 4	AIC	4	0.08	0.15	0.67
	BIC	4	0.08	0.15	0.67

Table 5: Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test results, as measured by p-values, for the daily EUA (panel a) and TTF (panel b) price series for the Full Sample and Samples 1 to 4 as previously specified. The model specification are trend=false and trend=true referring to Equation 6 where $\delta = 0$ for the trend=false specification. P-values above 0.1 are reported as 0.1 and p-values below 0.01 are reported as 0.01. P-values above 5% are highlighted in bold. $\sqrt{T}$ for the full sample and samples one to four are 38, 22, 22, 22, and 16 respectively.

(a) EUA prices

		Lags		
	Trend	1	10	$\sqrt{T}$
Full sample	false	0.010	0.010	0.010
	true	0.010	0.010	0.010
Sample 1	false	0.010	0.010	0.010
	true	0.010	**0.068**	**0.100**
Sample 2	false	0.010	0.010	**0.059**
	true	0.010	0.010	0.010
Sample 3	false	0.010	0.010	0.010
	true	0.010	0.010	0.010
Sample 4	false	0.010	0.010	0.100
	true	0.010	0.010	**0.055**

(b) TTF prices

		Lags		
	Trend	1	10	$\sqrt{T}$
Full sample	false	0.010	0.010	0.049
	true	0.010	0.010	0.010
Sample 1	false	0.010	0.010	**0.100**
	true	0.010	0.010	**0.064**
Sample 2	false	0.010	0.010	0.010
	true	0.010	0.010	**0.100**
Sample 3	false	0.010	0.010	0.010
	true	0.010	0.010	0.010
Sample 4	false	0.010	0.010	0.049
	true	0.010	0.010	0.021

Table 6: Johansen test results expressed in p-values for cointegration between the EUA and TTF price series for model specification H1*,H1, and H* for the Full Sample and Samples 1 to 4 as previously specified. The underlying VEC(q) models for the three model specifications are described by Equations 26,27, and 28, respectively. The test results are shown for the null hypotheses of rank = 0 and that the rank = 1 under both the trace test and the maximum eigenvalue test.

		trace test		maximum eigenvalue test	
		rank = 0	rank = 1	rank = 0	rank = 1
Full sample	H1*	0.25	0.87	0.12	0.87
	H1	0.15	0.77	0.11	0.77
	H*	0.48	0.71	0.44	0.71
Sample 1	H1*	0.58	0.45	0.72	0.45
	H1	0.49	0.07	0.77	0.07
	H*	0.20	0.65	0.15	0.65
Sample 2	H1*	0.62	0.88	0.48	0.88
	H1	0.30	0.33	0.34	0.33
	H*	0.12	0.49	0.12	0.49
Sample 3	H1*	0.38	0.42	0.49	0.42
	H1	0.19	0.11	0.35	0.11
	H*	0.49	0.35	0.73	0.35
Sample 4	H1*	0.07	0.52	0.06	0.52
	H1	0.02	0.08	0.04	0.08
	H*	0.37	0.88	0.19	0.88

Table 7: Likelihood ratio test results expressed in p-values for pairwise tests for H1*, H1, and H* test set ups for Sample 4. P-values below 0.05 suggest the restricted model is rejected in favor of the unrestricted model and p-values above 0.05 suggest the restricted model cannot be rejected at the 5% significance level.

	H1*	H1	H*
H1*	-	0.39	0.39
H1	-	-	0.96
H*	-	-	-

Table 8: Results of the parameter estimation of the VEC model for the cointegration matrix C.

	coefficient estimate	standard error	t-stat	p-value
$C(1,1)$	-0.0609	0.0196	-3.1098	0.0019
$C(2,1)$	-0.0614	0.0185	-3.3260	0.0009
$C(1,2)$	-0.0374	0.0120	-3.1098	0.0019
$C(2,2)$	-0.0377	0.0113	-3.3260	0.0009

Table 9: Results of the adjustment speed test within the Johansen framework for Sample 4 and model specification H1*. The EUA row shows the test results for the null hypothesis including $A_r^1 = \begin{bmatrix} 0 \\ -0.1292 \end{bmatrix}$ in the restricted model and the TTF row shows results for the null hypothesis including $A_r^2 = \begin{bmatrix} -0.1281 \\ 0 \end{bmatrix}$ in the restricted model. P-values below 0.05 suggest the restricted model is rejected in favor of the unrestricted model and p-values above 0.05 suggest the restricted model cannot be rejected at the 5% significance level.

	test statistic	critical value	p-value
EUA	7.232	3.841	0.007
TTF	8.244	3.841	0.004

Table 10: Results of the component share and information share measures of price discovery. Coefficient estimates and variance covariance matrix of the residuals are derived from the VEC(q) model estimated using the Sample 4 data and the H1* Johansen form.

	CFW	IS_{lower}	IS_{upper}	$IS_{average}$
EUA	114.02	0.382	0.687	0.534
TTF	-113.02	0.329	0.600	0.465