UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Price and Volatility Prediction in the EU ETS Market

Bachelor Thesis in Financial Economics, 15 hp
Department of Economics
Autumn 2013

Authors:
Gustav Ljungqvist
David Palmqvist

Supervisor:
Mohamed-Reda Moursli

Abstract

In this thesis we examine return and volatility predictability of continuous futures contracts within the European Union Emissions Trading System (EU ETS). The market has been active for nine years and we examine whether it is more mature now compared to a few years ago when most existing research was carried out. We find that autoregressive terms are now significantly weaker compared to during the first phase of the ETS, which is seen as a sign that the market has become more efficient. As heteroskedasticity is observed, GARCH models are used to model and predict volatility. To predict returns, we find that using exogenous inputs, in the form of electricity, coal, Brent oil and gas prices, yield better results than using autoregressive terms of the emission allowance data. Based on the results, we suggest that exogenous variables may be used to predict the returns of carbon futures.

Acknowledgments

We would like to express our appreciation to our supervisor Mohamed-Reda Moursli for his insightful help and constructive suggestions. We would also like to thank our friends Christian and Johan, for the many rewarding discussions and the moral support over the course of writing this thesis.

Contents

1 Introduction

The EU emissions trading system (EU ETS) was launched in 2005 to combat climate change mostly due to large emissions of carbon dioxide (CO_2). Several other emission systems provided experience prior to the launch of the EU ETS, such as the UK Emissions Trading Scheme (UK ETS), the Danish CO_2 trading program and the US sulfur dioxide (SO_2) emissions trading system. The EU ETS shows many similarities to the US system, there are however several key differences. The ETS incorporates a much larger set of companies, the emission reduction rate is lower, and the value of the traded allowances of the ETS is about 8 times larger than that of the US SO_2 traded allowances. (Ellerman and Buchner, 2007).

As of 2013, more than 11 000 power generation and manufacturing firms are part of the ETS. It covers about 45% of the participating countries greenhouse gas emissions. The participants are spread across the 28 EU countries as well as Iceland, Lichtenstein and Norway. As a cap and trade system the ETS sets a cap on the amount of greenhouse gases that can be emitted. The overall goal is to lower greenhouse gases, this is done by annually reducing the limit of allowances available on the market. From 2005 to 2020 the allowances will be reduced by 21%. Some allowances are allocated for free, while the rest are being auctioned. In 2013 more than 40% of the allowances were auctioned, and by 2020 the allocation of free emission allowances is supposed to stop, and the allocation process will instead be based solely on auctioning. (European Commission, 2013).

The allowances can also be traded on the secondary market directly between participants, through a broker or on an exchange. The largest part of the traders are companies that are obligated to own allowances because of their emissions, however a considerable portion of the trading is driven by hedging, portfolio adjustments, profit taking and arbitrage (Kossoy and Guigon, 2012).

Since the ETS is an immature market there has been many price jumps, especially during the first two years the price fell sharply on several occasions. The drastic changes in price was the result of over-allocation (Ellerman and Buchner, 2007). The 2005 to 2007 period was the first phase of the ETS. The second trading period, or phase two, lasted from 2008 to 2012. During the second period the total amount of allowances was reduced by 6.5%, but because of the financial crisis the price of the emission allowances was still lower than expected. The third phase started in 2013 and ends at the end of year 2020. A big change from previous phases is the introduction of an overall EU cap, instead of

individual country caps. The annual reduction rate during phase three is set to 1.74% (European Commission, 2013).

Prediction of returns is an intriguing subject in a young and growing market such as the ETS. As the amount of instruments (e.g. spots and different types of forward and futures contracts) increases, traders are interested not only in long-term price behavior, but also short-term price dynamics. Most existing literature on predicting the European Union allowance (EUA) price and volatility use data from the first phase of the EU ETS. Since the market was very young at the time of their research, one might suspect that it has stabilized by now. Moreover, structural changes have occurred, and the trading volume has significantly increased since the first phase.[1] Furthermore, as seen in the following chapter, several papers have been published on how closely related the ETS market price is to other market fundamentals, such as various energy prices. However, literature using exogenous factors to predict the EUA price is sparse. Hence we state two hypotheses:

- *The EU ETS market has come closer to being an efficient market and therefore the potential of prediction has decreased.*

- *Including exogenous variables, such as energy related commodity prices, increases the predictive power of our models and lead to smaller errors when predicting.*

The first hypothesis will be investigated by constructing ARMA-GARCH models based on a sample covering the second and the start of the third phase of the ETS. If our hypothesis is true, trends might not be as prominent as earlier research, based mostly on data from the first phase, has shown. In this case we will most likely find overall less significant parameters, and probably smaller autoregressive coefficients in our models, as Fama (1970) states that lack of autocorrelation indicates an efficient market. The exogenous variables tested will be electricity, Brent oil, coal and natural gas prices, as well as a paper price index and a stock index.

Indeed our results strengthen both of our hypotheses. We find only very small autoregressive terms when modeling the EUA returns. Moreover, we find that the prices of electricity, Brent oil, coal and natural gas can be used to predict the EUA price. Doing so decreases both the mean squared and mean absolute errors (MSE and MAE) of the predictions, compared to using only the past EUA returns.

---

[1]In 2011, the total traded volume of EUAs were 7.9 billion tonnes CO_2 or 148 billion USD (Kossoy and Guigon, 2012), compared to 0.3 billion tonnes CO_2 or 8.2 billion USD for 2005 (Capoor and Ambrosi, 2006).

2 Related literature

The first research on the ETS, carried out around and just after its launch in 2005 focused on finding determinants for the EUA price. Pioneering work was done by Christiansen et al. (2005) who identified three key drivers for the market price in the ETS, these three being policy and regulatory issues, market fundamentals, and technical indicators. Following this, Mansanet-Bataller et al. (2007) were the first to investigate econometrically the relationships between energy markets and the EUA price. They found that the most emission intensive energy variables, i.e. coal, Brent and to some extent natural gas, are the most important ones in the determination of EUA returns. Their work was extended by Alberola et al. (2008) who showed that the EUA price is related to the prices of Brent oil, natural gas, coal, electricity, and energy spreads[2]. They also investigated the effects of temperature on the EUA price, and the consequences of two structural breaks which occurred during the first phase of the ETS.

Concerning technical indicators, Paolella and Taschini (2008) used an AR(1)-GARCH(1,1) model to capture heteroskedasticity in the EUA returns. They also look at the US SO_2 market, where they find an "extremely mild" autoregressive term. Benz and Trück (2009) also worked with an AR(1)-GARCH(1,1) model, and found a large and significant autoregressive term for the EUA returns. They also employed a Markov regime-switching model, which performed slightly better than the AR-GARCH model when doing out-of-sample predictions. Other mean estimating models than the AR(1) are rarely used, exceptions include Alberola and Chevallier (2009) who use an ARMA(1,1) model.

As the first phase of the ETS had ended, Daskalakis et al. (2009) did explicit modeling on the consequences of the prohibition on inter-phase banking of allowances that was in force between the first and second phases of the ETS. They claim that this prohibition "may have an adverse effect on market liquidity and efficiency" (p. 1231), and indeed it is believed to have been a major reason for the price crash in 2007 (Ellerman and Joskow, 2008). For the second and all subsequent trading phases, unrestricted inter-phase banking has been allowed.

Combining exogenous variables, such as coal, Brent and natural gas, with structural models such as the GARCH model was done by Chevallier (2009). He employed a TGARCH(1,1) model for the variance, and used several macroeconomic and energy variables for the mean equation. However, no actual forecasting was carried out in his paper.

---

[2]Energy spreads are measures of the difference between the market price of electricity and its cost of production, when produced using either coal (dark spread) or natural gas (spark spread).

Using data from the second phase of the ETS, Chevallier (2011) found that the EU industrial production index does impact the EUA futures price. He also suggested the emission allowance market to be related to Brent oil and natural gas, but not coal. However, Mansanet-Bataller et al. (2011) suggested that the EUA futures price is affected by Brent oil, natural gas and coal. Aatola et al. (2013) further investigated determinants of the EUA price. They found that not only energy variables, with electricity price being the most important one, but also factors such as a stock index as well as paper and mineral price indices impacts the price of an emission allowance. Recently, Byun and Cho (2013) used several ARMA-GARCHX models, that is, with exogenous inputs (electricity, oil, coal and gas prices) in the variance equation. They analyzed forecasting and concluded that indeed electricity, coal and Brent oil prices can be used advantageously to forecast the volatility of emission allowance prices. They also find that ARMA-GARCH models considering more than one lag can be rejected according to Bayesian information criteria, which is a finding in line with previous research.

3 Methodology

In this section we explain the steps that will be carried out to obtain a model. As mentioned in section 2, AR-GARCH or ARMA-GARCH models are often used in modeling and forecasting EUA prices. The GARCH part models conditional heteroskedasticity, which is useful in financial time series since volatility clustering is often observed. For the mean equation, an ARMAX model allows for autoregressive terms, moving average terms, and also exogenous inputs. Hence, for this thesis we examine ARMAX-GARCH models, which are specified in equations 1-3. We first try to find the best ARMA-GARCH model, and then add exogenous terms.

First we need to check the data sets for stationarity, as carried out according to section 3.1. If we can not reject non-stationarity, ARMAX-GARCH models will not be suitable. If we have stationary time series, we construct a few ARMA models and perform ARCH LM tests, as described in section 3.2, and add ARCH and GARCH terms accordingly. When we have our ARMA-GARCH models, we estimate parameters and evaluate the models based on their Bayesian information criteria values (BIC-values) as described in section 3.3.

When we have found our optimal ARMA-GARCH model, we introduce exogenous variables. We start with several different factors and evaluate them as specified in section 3.4. Subsequently we construct models with all possible constellations of the exogenous variables that we have selected. We look not only at their BIC-values, but also at likelihood ratio tests (see section 3.5), where we test whether the exogenous variables in each model add any explanatory power compared to the benchmark ARMA-GARCH model. We also perform Ljung-Box Q-tests as described in section 3.6. Finally, we do out-of-sample predicting with our ARMAX-GARCH models, and evaluate their performances, as described in section 3.7.

When we use ARMAX models we only look at one lag for the exogenous variables. Hence we specify our models as $ARMAX(r, m, b)$ models, where r denotes the number of autoregressive terms, m the number of moving average terms, and b the amount of exogenous time series. The mathematical formulation for an $ARMAX(r, m, b)$-$GARCH(p, q)$

is

$$y_t = c + \sum_{i=1}^{r} \phi_i y_{t-i} + \sum_{j=1}^{m} \theta_j \epsilon_{t-j} + \sum_{k=1}^{b} \eta_k d_{k,t-1} + \epsilon_t \tag{1}$$

$$\epsilon_t = u_t \sigma_t \tag{2}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{3}$$

Here, equation 1 is the mean equation, equation 2 describes the residuals, and equation 3 is the conditional variance equation. y_t is the return of the EUA price at time t, c is a constant, ϕ_i is the AR parameter at lag i, θ_j is the MA parameter at lag j, η_k is the parameter for the k:th exogenous dataset and $d_{k,t-1}$ is the return of the exogenous variable k at time $t - 1$. ϵ_t is the residual at time t, which is the product of u_t, assumed to be independent and identically distributed, zero mean and with unit variance and σ_t, which is described by equation 3. Here α_0, α_i and β_j are real constants, where α_i is the ARCH parameter and β_j the GARCH parameter at lag i and j respectively.

3.1 Stationarity

To model our time series as an ARMA-GARCH model the series should be stationary. The Augmented Dickey-Fuller (ADF) test checks for a unit root in the time series. If a time series has a unit root it is non-stationary. We conduct the test without including an intercept or trend term, which means that if the null hypothesis is true, the time series is a random walk. The alternative hypothesis is that the series is stationary. The number of lags is chosen by minimizing the Akaike information criterion (AIC), which is one method suggested by, for example, Hall (1994). The other test we use is the KPSS test (Kwiatkowski et al., 1992). Since we are not including a trend term, the null hypothesis of the KPSS test is that the series is level stationary. The alternative hypothesis is that it has a unit root. When we choose the number of lags we follow one of the suggestions from Schwert (1989), who does a simulation study of unit root tests. One of the lag lengths he advocates, and the one we use, is defined as the integer value of $4(n/100)^{1/4}$, where n is the number of observations.

3.2 Testing for ARCH effects

Before we construct a volatility model we check the residuals of the mean equation for ARCH effects. When looking for ARCH effects we use the Lagrange multiplier test of Engle (1982). The test is applied to the squared residuals of the mean equation, and

the null hypothesis is that there is no autocorrelation among them. The alternative hypothesis is autocorrelation in the squared residuals. If the null hypothesis is rejected we can assume that there are ARCH effects present. We use 12 as the number of lags as is done by Tsay (2010), among others.

3.3 Likelihood function and the Bayesian information criterion

When estimating parameters in a model, one usually seeks to maximize the likelihood function. It is often more convenient to work with the natural logarithm of the function, i.e. the log-likelihood function (LLF), which obviously takes its optimal value for the same parameters. The LLF will never decrease when more parameters are added, hence selecting models solely based on their maximized LLF value might lead to an unnecessarily complicated model. Instead the Bayesian information criterion, developed by Schwarz (1978), is often used for model selection. The formula is

$$BIC = -2\hat{L} + k\ln(n)$$

where $\hat{L}$ is the maximized LLF value, k is the number of parameters, and n is the amount of observations. A lower BIC value indicates a better model. The BIC is frequently used in time series or linear regression modeling, and the literature on EUA return and volatility prediction is no exception.[3]

3.4 Choosing exogenous variables

For the ARMAX models considered, we use as exogenous variables the daily returns (lagged by one business day) of other commodities as well as macroeconomic factors. We start by looking at six factors found to be important in determining the EUA price in earlier studies. Those factors are electricity, coal and gas prices, a European stock market index (we use STOXX Europe 600) and a price index for paper products, all found significant by Aatola et al. (2013). We also use the Brent oil price, which as mentioned in section 2 has been found significant in numerous other studies.

Since many companies covered by the ETS are producing electricity or paper, these prices should theoretically impact the price of an EUA. For example, when the electricity price is high, power plants want to produce a lot of electricity, for which they will need more EUAs. The demand for emission allowances increases, and subsequently so does

---

[3]Research using BIC includes for example Paolella and Taschini (2008), Daskalakis et al. (2009), Benz and Trück (2009) and Byun and Cho (2013).

the price. The coal price is expected to have a negative impact on the EUA price, as coal is a very emission intensive fuel. When coal is expensive, power plants will be less inclined to use coal, and hence not need as many emission allowances. The opposite goes for natural gas, as it is a fuel emitting less CO_2 than coal. Therefore when natural gas is expensive, power plants will likely switch to coal instead of natural gas, increasing the need for emission allowances. The stock index will theoretically have a positive effect, as high economic activity will lead to an increase in demand of EUAs. Concerning the Brent oil price effect, most studies find that the oil price has a positive influence on the EUA price. It remains unclear whether this should be attributed to a fuel switching effect, as oil is a less carbon emitting fuel than coal, or rather to the correlation between the oil price and economic activity (Rickels et al., 2010). A summary of these theory based impacts can be found in table 1.

Variable	Description	Impact
Electricity	End product price	$+$
Paper	End product price	$+$
Gas	Less emitting input price	$+$
Stock index	Economic activity	$+$
Brent oil	Less emitting input price and/or economic activity	$+$
Coal	More emitting input price	$-$

Table 1: Summary of the theory based impacts of exogenous factors on the EUA price.

Note that we in our models use lagged returns of exogenous variables, so the signs of parameters might not be consistent with the theoretical signs. Previous research considers mostly non-lagged modeling, to determine price drivers. Hence, we do not know if the lagged values of our exogenous variables are relevant in predicting the EUA returns. To find out which ones that are relevant, we run the best ARMA-GARCH model found with one exogenous variable added to the mean equation, i.e. an ARMAX$(r, m, 1)$-GARCH(p, q) model. We do this for each variable one by one, and reject the factors that have a parameter which is not significant at the 5% level. The variables that pass this test will also be checked for multicollinearity. Following Chevallier (2009), we examine the cross-correlations between the variables, and check if any two variables have a correlation larger than 0.6 which is suggested to be a breaking point in his paper. If so, we remove the one of them that has larger cross-correlations overall. Furthermore, also following Chevallier (2009), we examine the variance inflation factor (VIF) for each variable. The VIF, proposed by Marquardt (1970), is found by running OLS regressions

for each variable as a function of the other variables, and it is specified as follows

$$\text{VIF}_i = \frac{1}{1 - R_i^2}$$

where VIF_i is the variance inflation factor for the i:th variable, and R_i^2 is the R squared value for the corresponding regression. The square root of the VIF_i can be interpreted as follows: It reveals how much larger the standard error is compared to what it would have been if variable i was uncorrelated with the other variables. If VIF_i is around or above 10 then multicollinearity is high. If so, we remove the i:th variable.

3.5 Likelihood ratio test

When choosing our ARMAX model we take advantage of the likelihood ratio test to see if our models add any explanatory power compared to the benchmark model. When adding more parameters, as we do when allowing for exogenous inputs, the likelihood will always increase. With this test we can determine if the added explanatory power is significant or not. The null hypothesis of the test is that the alternative model does not fit the data better. The formula for the test is

$$D = -2\hat{L}_0 + 2\hat{L}_a$$

where $\hat{L}_0$ and $\hat{L}_a$ are the maximized log-likelihood function values for the benchmark and alternative models respectively. The test statistic D has a χ^2-distribution with degrees of freedom equal to the difference in amount of parameters between the two models.

3.6 Ljung-Box Q-test

We use the Ljung-Box Q-test to test for autocorrelation of the squared standardized residuals of a fitted model. The null hypothesis is that the residuals show no autocorrelation, while the alternative is that there is correlation among the residuals. We use the test to see if our models are adequately fitted or not. If they are, there should be no autocorrelation present. The test statistic is defined as

$$Q(l) = n(n+2) \sum_{k=1}^{l} \frac{\hat{\rho}_k^2}{n-k}$$

where n is the number of observations, l is the number of lags to be tested and $\hat{\rho}_k$ is the sample autocorrelation function. According to Tsay (2010), simulation studies suggests using $\ln(n)$ as the number of lags. It is also suggested to use several different lags when

testing, we use $\ln(n)$, $\ln(n) + 5$ and $\ln(n) + 10$.

3.7 Prediction

We estimate our model parameters from a data sample containing a large amount of daily observations, the in-sample. After obtaining the parameter values, we test the predictive ability of our models on a smaller sample, the out-of-sample data. We use a rolling window technique to estimate models, so we only predict one day at a time, and then re-estimate the parameters. The length of the sample used to estimate parameters is kept constant, i.e. the start date and end date successively increases with one observation.[4]

An important step in predicting lies in the evaluation of the predictions. We use a few different performance measures to assess prediction ability in our forecasts. The MSE is often used in this context. A downside of this measure is that outliers have a heavy impact on the result since it uses the squared errors. The outlier problem is not as prominent when using MAE for evaluation. The MSE and MAE are calculated with the differences of the predicted returns and the actual returns, the formulas can be seen in appendix A.2.

---

[4]This technique is very common and in this field it is used by for example Paolella and Taschini (2008) and Byun and Cho (2013).

4 Data

We use daily data from the ICE European Climate Exchange (ICE ECX) emission allowance continuous futures contract. The continuous future uses a rolling method where a new future is added to the series when the last one has matured. It has become common practice to use this kind of derivative for academic studies (Chevallier, 2010). Trück et al. (2012) find that the price behavior of the EUAs spot and futures market is very different from other commodities. We do not use the spot prices, instead we focus entirely on the price of futures contracts. They further state that market participants have a tendency to hold long futures positions instead of the spot, and a large majority of the number of allowances traded comes from traded futures, while the spot trading contribution is much smaller. The ICE ECX is by far the most liquid market place for EUA futures contracts (IntercontinentalExchange Group, 2011). As the vast majority of existing literature on the subject, we use daily data to be able to compare our results to previous research. The in-sample dataset runs from 2009-11-04 to 2013-07-01 (953 observations), and we have an out-of-sample period of 2013-07-02 to 2013-11-04 (90 observations) on which we test the performance of our models. The source of our data is Thomson Reuters Datastream, unless otherwise stated. As we are interested in capturing the growth rate of the dependent variable, we use logreturns, that is

$$y_t = \ln(s_t) - \ln(s_{t-1})$$

where s_t is the price at time t. Both the prices and logreturns are plotted in figure 1. The descriptive statistics of this data can be seen in table 2.

Dataset	Size	Mean	Min	Max	Std. Dev.	Skew.	Kurt.
In-sample	953	-0.0013	-0.4314	0.2453	0.0365	-1.2536	27.6589
Out-of-sample	90	0.0008	-0.1134	0.0893	0.0320	0.1795	4.9608

Table 2: Summary statistics for the EUA futures logreturns y_t. *Std. Dev.* refers to standard deviation, *Skew.* to the skewness, and *Kurt.* to the kurtosis.

Since the kurtosis of the in-sample data is as large as 27.6589, we assume that our returns are t-distributed rather than normally distributed. An illustration of this is given in figure 2, which also indicates that a t-distribution indeed fits the data better. A GARCH model in combination with a t-distribution for error terms was first used by Bollerslev (1987). Note that when considering the t-distribution we have to estimate one additional parameter for each model, the degrees of freedom, denoted by ν.
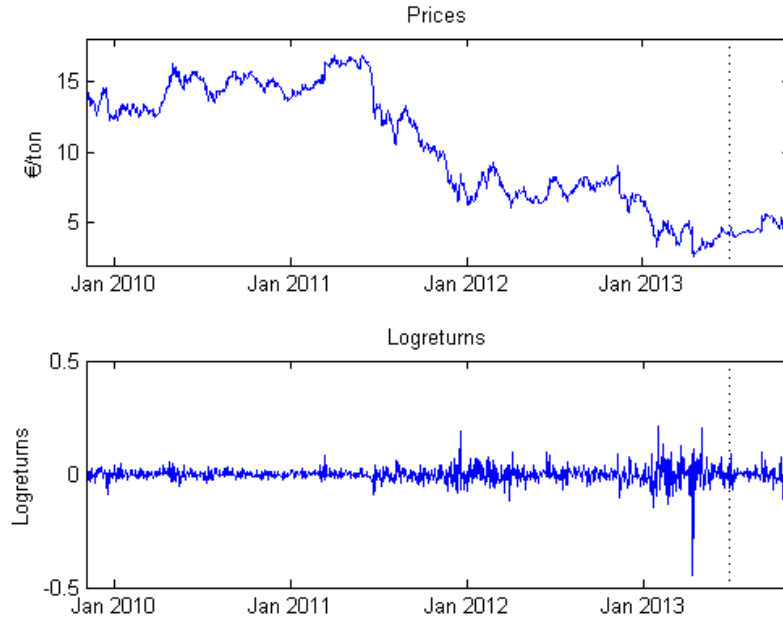
Figure 1: The price and logreturns of the EUA continuous futures contract, from November 2009 to November 2013. The dashed line indicates where the out-of-sample period starts.
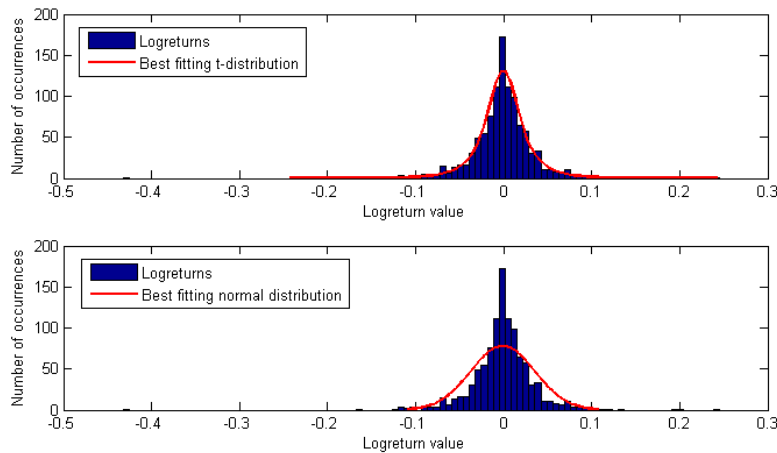
Figure 2: Histograms of our in-sample data, together with the best fitting t-distribution (above) and normal distribution (below).

4.1 Exogenous data

As mentioned in section 3.4 we start by looking at six different factors that are related to the price of an EUA. These are the returns of electricity, oil, coal and gas prices, as well as a paper price index and a stock index. Here, the electricity variable is the EEX (European Energy Exchange) yearly baseload continuous futures contract. The Brent oil variable is the ICE Brent crude oil continuous futures contract, the coal variable is the

EEX coal ARA month continuous futures contract and the gas variable is the ICE UK Natural Gas 1-month continuous futures contract. Paper is the FOEX PIX EU Paper A4 B-copy index, and the stock index is the STOXX Europe 600. All of these run on the same interval as the EUA data. The futures that are not expressed in EUR have been converted using the exchange rates from the European Central Bank (ECB).[5] Summary statistics for the exogenous data can be found in appendix C.2.

4.2   EUA data treatment

The autocorrelation function of the in-sample data tells us about the potential of AR models to predict future returns. It is plotted in figure 3. The most important thing to notice is that the autocorrelation coefficient with lag 1 is very small, 0.03, and not statistically significant. Also, there seems to be no clear pattern in which lags that have significant autocorrelations. To investigate this further, we look at the autocorrelation function when some outliers are removed, to see whether the autocorrelation coefficients keep the same values or not. As seen in figure 4, the coefficients dramatically changes when only the three most extreme returns are removed from the set, which indicates poor robustness in the autocorrelations. We know that for a randomly distributed variable, the autocorrelations should be between $\pm z_{1-\alpha/2}\sigma$ with significance level α, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$:th quantile of the normal distribution. We note that for $\alpha = 0.05$, lag 2 violates this in all three of our cases, which indicates that the statistical significance of this particular autocorrelation is robust.
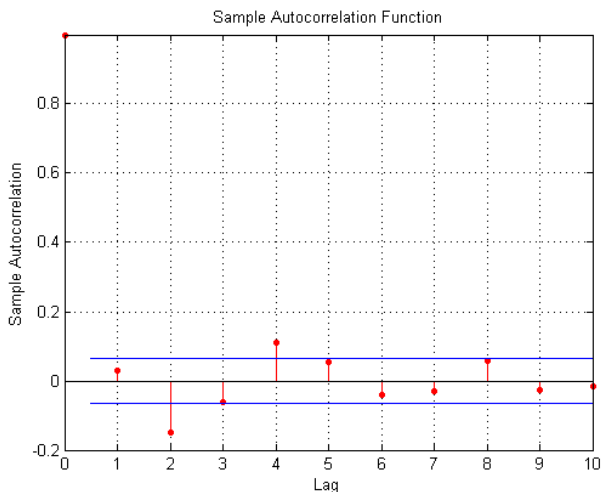
Figure 3: The in-sample autocorrelation function with 95% confidence bounds.
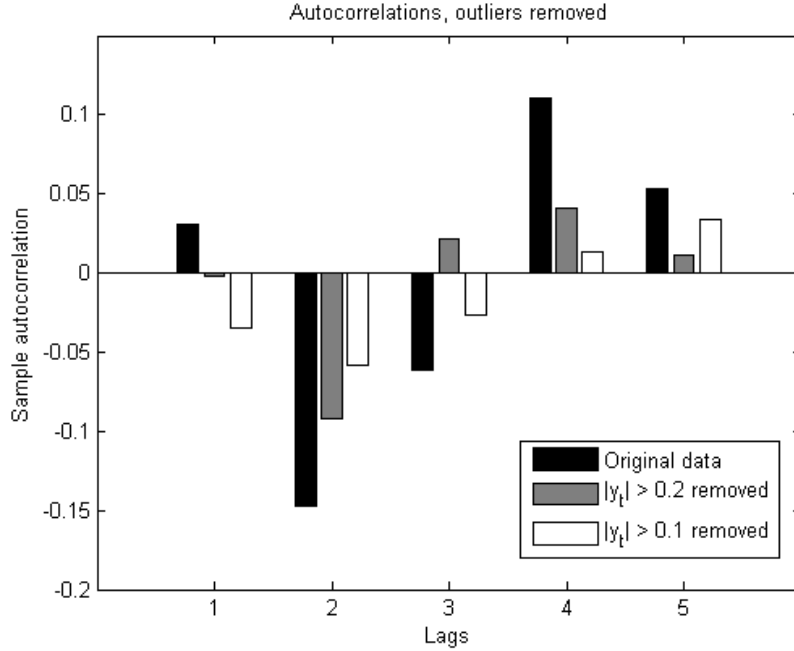
Figure 4: The in-sample autocorrelation function with some outliers removed. Only 3 observations meet $|y_t| > 0.2$, but still the autocorrelation coefficients are very different from the original case. 18 observations meet $|y_t| > 0.1$.

4.2.1 Stationarity

A visual inspection of figure 1 suggests that the logarithmic return series is stationary, but as specified in section 3.1 we use a couple of different tests to make sure that it is. After performing both the ADF test and the KPSS test, we can conclude that the time series used are stationary. The results from the tests can be seen in appendix C.3. The ADF test reports a p-value of less than 0.01 for all return series, which means we reject the null hypothesis of a unit root. The reported p-values of the KPSS test support the ADF test, i.e. we cannot reject the null hypothesis of stationarity at the 5% significance level.

5 Results

Here we construct various models by estimating parameters using the in-sample data. From the autocorrelation function and the robustness of it examined earlier, we know that the potential of AR and MA models are slim. We evaluate a few of those models, and the results, seen in tables 3 and 4 and discussed below, confirm that they are not suitable here. When applying the ARCH LM test to the different ARMA models it is also apparent that there is conditional heteroskedasticity present. From the table in appendix C.4 we see that we can reject the null hypothesis for all ARMA models at the 1% level at 12 lags. Note that we have also evaluated the models using normally distributed errors,

Model	k	$\hat{L}$	BIC
AR(0)	3	1992.7	-3964.8
AR(1)	4	1992.7	-3958.0
AR(2)	5	1999.8	-3965.4
AR(3)	6	2001.1	-3961.1
AR(4)	7	2002.9	-3957.7
MA(1)	4	1992.7	-3958.0
MA(2)	5	1999.1	-3963.9
ARMA(1,1)	5	1994.8	-3955.3
ARMA(2,2)	7	2003.1	-3958.7
GARCH(1,1)	5	2140.8	-4247.3
AR(1)-GARCH(1,1)	6	2140.9	-4240.6
MA(1)-GARCH(1,1)	6	2140.9	-4240.7
AR(2)-GARCH(1,1)	7	2141.6	-4235.2
MA(2)-GARCH(1,1)	7	2141.6	-4235.2
ARMA(1,1)-GARCH(1,1)	7	2142.3	-4236.7
ARMA(2,2)-GARCH(1,1)	9	2144.4	-4226.1
GARCH(2,2)	7	2141.8	-4235.6
GJR-GARCH(1,1)	6	2143.5	-4245.9

Table 3: Basic performance evaluation of the estimated models. k is the number of estimated parameters, $\hat{L}$ is the maximized value of the log-likelihood objective function, and BIC is the Bayesian information criterion.

as opposed to the t-distributed errors we use here and for the remainder of this paper, and these results can be seen in appendix C.5.

In table 3 we have listed a number of evaluated models. We evaluate their performance mainly based on their Bayesian information criterion. Looking at the mean equation models exclusively, we see that the best performing model is the AR(2) model, but it is only slightly better than the AR(0)-model, which suggests that the autoregressive terms

Model	AR(0)	AR(1)	AR(2)
c	-0.0002	-0.0002	-0.0002
	(0.0008)	(0.0008)	(0.0008)
ϕ_1		-0.0006	-0.0214
		(0.0217)	(0.0213)
ϕ_2			-0.1306***
			(0.0198)
σ^2	0.0022**	0.0022**	0.0022**
	(0.0011)	(0.0011)	(0.0011)
ν	2.3940***	2.3935***	2.3982***
	(0.2573)	(0.2572)	(0.2582)

Table 4: Parameter values of AR models, according to equations 1-3. Standard deviations are in parentheses. *** indicates significance at 1% level, ** at 5% and * at 10%.

might not be very strong in these models. This is in strong contrast to the AR(1) model of Benz and Trück (2009), which had a statistically significant ϕ_1 with a value of 0.2122. As stated earlier, they looked at the first phase of the ETS, and we look at the second and start of the third phase. We interpret this as an indication that trends are less prominent and possibly that the market has matured since the first phase, which strengthens our hypothesis of a more efficient market. One reason for the large difference could be the structural change of allowing inter-phase banking of allowances that occurred at the start of the second phase, which has been mentioned in section 2. Another likely reason is the dramatic increase in trading volume, which has been mentioned in section 1.

The MA models are worse than the AR models of the same lags, and the ARMA models have even worse BIC. Several higher order AR models have been tested but according to the BIC, they were considered far worse than the first two. Further down the table, we see that when allowing for heteroskedasticity, we achieve far higher BIC values, however, when considering heteroskedastic models, ARMA terms only lowers the BIC value. We also evaluate one specification of the GJR-GARCH model[6], which Byun and Cho (2013) found to perform well. The best BIC value is achieved for a simple GARCH(1,1) model, and hence we rate it as the best fitting model.

Table 4 showcases parameter values of the three best AR models. Notably, among the AR parameters only the AR(2) parameter ϕ_2 is statistically significant. Proceeding to the conditional heteroskedasticity models in table 5, we note that no AR parameters at

---

[6]The GJR-GARCH is named after Glosten, Jagannathan, and Runkle who first suggested it. It differs from the ordinary GARCH model in that it also models asymmetry in the ARCH process (Glosten et al., 1993).

all are significant. On the other hand, GARCH parameters are significant all the way down to the 1% level. Hence, we only use the GARCH(1,1) model without AR terms in the next section.

Model	GARCH(1,1)	AR(1)-GARCH(1,1)	AR(2)-GARCH(1,1)
c	-0.0002	-0.0002	-0.0002
	(0.0006)	(0.0006)	(0.0006)
ϕ_1		-0.0133	-0.0135
		(0.0324)	(0.0324)
ϕ_2			-0.0400
			(0.0308)
α_0	$6 \cdot 10^{-6}$**	$6 \cdot 10^{-6}$**	$6 \cdot 10^{-6}$**
	$(3 \cdot 10^{-6})$	$(3 \cdot 10^{-6})$	$(3 \cdot 10^{-6})$
α_1	0.1172***	0.1172***	0.1138***
	(0.0216)	(0.0216)	(0.0211)
β_1	0.8828***	0.8828***	0.8862***
	(0.0192)	(0.0192)	(0.0189)
ν	5.4060***	5.4020***	5.3092***
	(0.7998)	(0.8056)	(0.7894)

Table 5: Parameter values of GARCH and AR-GARCH models, according to equations 1-3. Standard deviations are in parentheses. *** indicates significance at 1% level, ** at 5% and * at 10%.

5.1 Exogenous inputs - ARMAX-GARCH

The ARMAX model allows for exogenous inputs in the mean equation. We have already seen that the optimal ARMA-GARCH model is ARMA(0,0)-GARCH(1,1). Therefore we will evaluate ARMAX(0,0,b)-GARCH(1,1) models when considering exogenous inputs.[7] Details regarding the exogenous variables have already been discussed in section 4.1. Now, as specified in section 3.4, we will run ARMAX(0,0,1)-GARCH(1,1) evaluations with each variable one by one. The results can be found in appendix C.6. The paper and stock indices have p-values higher than 0.05 and are therefore disregarded.

Before performing the multivariate regression that is the ARMAX equation with more than one exogenous variable, we need to check for multicollinearity. Again, as stated in section 3.4 we follow Chevallier (2009) and examine the cross-correlations between our energy variables. Results are found in appendix C.7, where we see that the highest in absolute value is around 0.4, which is substantially lower than the 0.6 suggested to be a

---

[7]Indeed, when evaluating for example ARMAX(1,0,b)-GARCH(1,1) models, we do not get statistical significance for the AR(1) term in any case.

breaking point in his paper. We also examine the variance inflation factor (VIF) for each variable, but no problematic collinearities are detected from these calculations either, as the highest VIF is the one for electricity with a value of 1.33, which is far from worrying. Complete results of these calculations can be found in appendix C.8.

With these four explanatory variables we construct one model for each combination of the variables, a total of 15 models. The model specifications can be seen in table 6. Note that this is inspired by Byun and Cho (2013), who uses exogenous variables for the variance equation (GARCHX). Table 7 contains model evaluation results for all of our

Model	Variables included			
Model 1	Electricity			
Model 2		Oil		
Model 3			Coal	
Model 4				Gas
Model 5	Electricity	Oil		
Model 6	Electricity		Coal	
Model 7	Electricity			Gas
Model 8		Oil	Coal	
Model 9		Oil		Gas
Model 10			Coal	Gas
Model 11	Electricity	Oil	Coal	
Model 12	Electricity	Oil		Gas
Model 13	Electricity		Coal	Gas
Model 14		Oil	Coal	Gas
Model 15	Electricity	Oil	Coal	Gas

Table 6: ARMAX-GARCH model specifications. Listed are the exogenous inputs considered in the mean equation.

ARMAX(0,0,b)-GARCH(1,1) models. We see that the one giving the best BIC value is model 14, which considers oil, coal and gas returns as exogenous variables. The p-values from the likelihood ratio (LR) test are generally very low, which indicates that adding energy variables adds a good amount of explanatory power to the benchmark GARCH(1,1) model. We can not reject the hypothesis that the GARCH(1,1) is better than or equal to the ARMAX models with only electricity and only oil as explanatory variables at the 1% level, but at the 5% level the null hypothesis can be rejected in all cases. Parameter values for all of these models when fitted to the in-sample data can be found in appendix C.9. We see that the coefficient of natural gas is statistically significant at the 1% level in all models. Brent oil and coal are in every case significant at the 5% but not always at the 1% level. Electricity is the least significant variable, as it is only statistically significant in model 1 and model 5.

Before we use our models for the predictions, we apply the Ljung-Box Q-test to confirm that the GARCH models we use adequately explains the variance in the return series. The null hypothesis of no autocorrelation among the squared standardized residuals can not be rejected, as can be seen in appendix C.10. This is in line with our expectations and means that the models are well fitted to the data.

Model	k	$\hat{L}$	BIC	LR-test p-value
GARCH(1,1)	5	2140.8	-4247.3	
Model 1	6	2143.6	-4246.0	$19.237 \cdot 10^{-3}$
Model 2	6	2143.2	-4245.3	$28.973 \cdot 10^{-3}$
Model 3	6	2146.8	-4252.5	$0.5217 \cdot 10^{-3}$
Model 4	6	2148.7	-4256.2	$0.0750 \cdot 10^{-3}$
Model 5	7	2147.3	-4246.6	$1.5285 \cdot 10^{-3}$
Model 6	7	2148.1	-4248.2	$0.6801 \cdot 10^{-3}$
Model 7	7	2149.6	-4251.1	$0.1600 \cdot 10^{-3}$
Model 8	7	2150.9	-4253.7	$0.0431 \cdot 10^{-3}$
Model 9	7	2153.1	-4258.1	$0.0048 \cdot 10^{-3}$
Model 10	7	2152.8	-4257.6	$0.0062 \cdot 10^{-3}$
Model 11	8	2151.7	-4248.5	$0.0737 \cdot 10^{-3}$
Model 12	8	2153.5	-4252.2	$0.0126 \cdot 10^{-3}$
Model 13	8	2152.9	-4250.9	$0.0240 \cdot 10^{-3}$
Model 14	8	2156.6	-4258.4	$0.0006 \cdot 10^{-3}$
Model 15	9	2156.6	-4251.6	$0.0023 \cdot 10^{-3}$

Table 7: ARMAX-GARCH model estimation results, compared to the optimal model without exogenous inputs, i.e. the GARCH(1,1). k is the number of estimated parameters, $\hat{L}$ is the maximized value of the log-likelihood objective function, and BIC is the Bayesian information criterion. LR-test p-value is the p-value obtained from a likelihood ratio test, where we reject the null hypothesis, i.e. that the benchmark model (GARCH(1,1)) is at least as good as the tested model, when the p-value is sufficiently low, as described in section 3.5.

5.2   Results from prediction

Results from predicting the standard deviations of the returns with the GARCH(1,1) model can be seen in figure 5. We see that volatility clustering seems to be featuring in the out-of-sample data as well, and the model partially captures it.

Prediction results for all of our ARMAX(0,0,b)-GARCH(1,1) models can be seen in table 8, along with the benchmark GARCH(1,1). Also included in the table are two AR-GARCH models to illustrate that they indeed perform worse than the ARMAX-GARCH
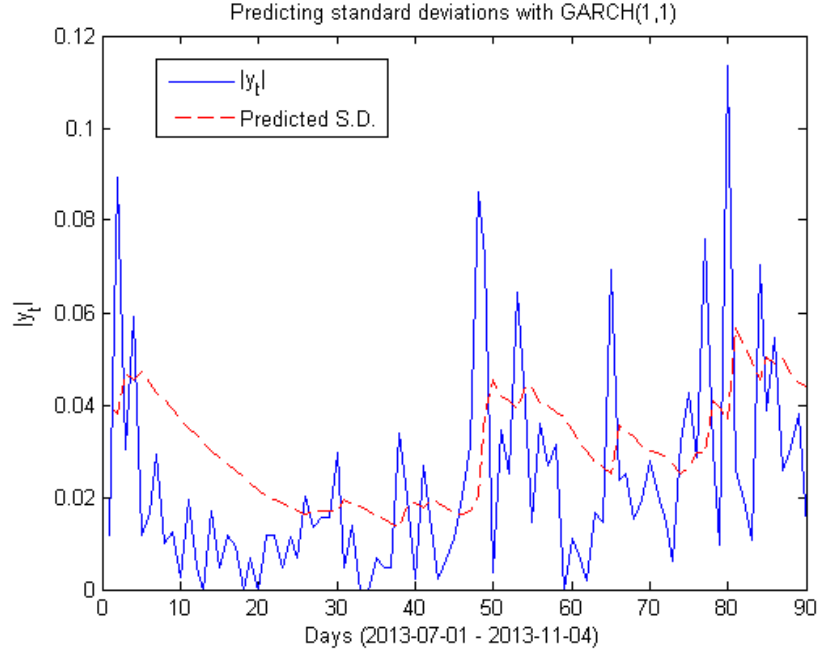
Figure 5: Results from predicting the standard deviations of the EUA returns with the GARCH(1,1) model, plotted along with the absolute values of the returns in the out-of-sample period.

models, just as most of the BIC values suggested. Moreover, the information criteria suggestion that most ARMAX-GARCH models should do better than the standard GARCH model is backed up by the actual prediction results. Even though the differences in the MSE and MAE between the models are quite small, the benchmark GARCH model performs worse in both error measures than each and every ARMAX-GARCH model. This confirms that incorporating exogenous inputs in the mean equation enhances prediction power. We note that models 1-4, i.e. the models that only consider one exogenous variable at a time, are overall performing quite badly (with the sole exception of the MSE of model 1), especially in mean absolute error. Taking both error measures in account, we see that the best two performances are given by model 5 (electricity and oil) and model 11 (electricity, oil and coal). In general, the models that include electricity as an exogenous variable perform better. Aatola et al. (2013) find that electricity is the market fundamental which has the biggest impact on the changes in the price of EUAs. Therefore it comes as no surprise that lagged returns of the electricity price are good for predicting.

| | MSE | | MAE | |
Model	Value	Rank	Value	Rank
GARCH(1,1)	$1.0118 \cdot 10^{-3}$	16	$2.2889 \cdot 10^{-2}$	16
AR(1)-GARCH(1,1)	$1.0124 \cdot 10^{-3}$	17	$2.2916 \cdot 10^{-2}$	18
AR(2)-GARCH(1,1)	$1.0202 \cdot 10^{-3}$	18	$2.2912 \cdot 10^{-2}$	17
Model 1	$1.0045 \cdot 10^{-3}$	1	$2.2757 \cdot 10^{-2}$	12
Model 2	$1.0117 \cdot 10^{-3}$	14	$2.2774 \cdot 10^{-2}$	14
Model 3	$1.0094 \cdot 10^{-3}$	9	$2.2770 \cdot 10^{-2}$	13
Model 4	$1.0118 \cdot 10^{-3}$	15	$2.2827 \cdot 10^{-2}$	15
Model 5	$1.0054 \cdot 10^{-3}$	2	$2.2633 \cdot 10^{-2}$	2
Model 6	$1.0066 \cdot 10^{-3}$	3	$2.2737 \cdot 10^{-2}$	9
Model 7	$1.0079 \cdot 10^{-3}$	5	$2.2754 \cdot 10^{-2}$	11
Model 8	$1.0098 \cdot 10^{-3}$	12	$2.2684 \cdot 10^{-2}$	6
Model 9	$1.0116 \cdot 10^{-3}$	13	$2.2724 \cdot 10^{-2}$	7
Model 10	$1.0092 \cdot 10^{-3}$	8	$2.2751 \cdot 10^{-2}$	10
Model 11	$1.0070 \cdot 10^{-3}$	4	$2.2621 \cdot 10^{-2}$	1
Model 12	$1.0088 \cdot 10^{-3}$	7	$2.2663 \cdot 10^{-2}$	3
Model 13	$1.0084 \cdot 10^{-3}$	6	$2.2731 \cdot 10^{-2}$	8
Model 14	$1.0095 \cdot 10^{-3}$	11	$2.2674 \cdot 10^{-2}$	5
Model 15	$1.0094 \cdot 10^{-3}$	10	$2.2670 \cdot 10^{-2}$	4

Table 8: Prediction results for all of our ARMAX-GARCH models along with GARCH(1,1) as well as AR(1)- and AR(2)-GARCH(1,1) results. MSE is short for mean squared error, and MAE is short for mean absolute error. We have also included the rank of each model when ordered both after their MSE and MAE values.

6   Conclusion

In this thesis we construct several models to model the return and the volatility of the most traded EUA futures contracts. We find that the logreturns of EUA continuous futures contracts is a stationary time series, which exhibits excess kurtosis and is t-distributed. We find that our $ARMA(r, m)$ coefficients are overall much lower than those of earlier papers. In particular our $AR(1)$ coefficient is close to zero and not statistically significant, which differs substantially from, among others, Benz and Trück (2009), who based their research on data from phase one. Based on the statement of Fama (1970) that a lack of autocorrelation indicates an efficient market, we conclude that the market has become more efficient.

We find conditional heteroskedasticity in the residuals, like in previous research, and we introduce a GARCH model to capture the volatility clustering, again in line with existing literature. We try a few different GARCH models but the best results are obtained from a simple GARCH(1,1) model. We note that an ARMA model for the mean equation adds no explanatory power according to the BIC. After the fitting of GARCH models, the Ljung-Box Q-test confirms that the ARCH effects have been removed since the squared standardized residuals are independent and identically distributed.

We allow for exogenous inputs in the mean equation of our model, and find that lagged returns of electricity, Brent oil, coal and natural gas prices are statistically significant. Lagged returns of a paper price index as well as a stock index are not significant, and are therefore disregarded. When allowing for exogenous inputs we find that the mean equation of an ARMAX-GARCH model is best modeled through an $ARMAX(0,0,b)$ model. The statistical significance of improved performance when incorporating exogenous variables in our model is confirmed with the likelihood ratio test. The Bayesian information criterion suggests that one-day lagged oil, coal and gas returns are the optimal variables. When using our models for predicting the MAE and MSE confirm that all $ARMAX(0,0,b)$-GARCH(1,1) models perform better than the models without exogenous variables. This supports our second hypothesis of improved predicting performance when allowing for exogenous inputs. Adding autoregressive terms to our mean equation does not improve the performance. The model with the best out-of-sample performance, considering both the MAE and the MSE, is either the model with electricity and oil as exogenous variables or the model with electricity, oil and coal as exogenous variables. In general, electricity adds the most predictive power out of our exogenous variables.

References

Aatola, P., M. Ollikainen, and A. Toppinen (2013). "Price Determination in the EU ETS market: Theory and Econometric Analysis with Market Fundamentals". In: *Energy Economics* 36, pp. 380–395.

Alberola, E. and J. Chevallier (2009). "European Carbon Prices and Banking Restrictions: Evidence from Phase I (2005-2007)". In: *The Energy Journal* 30.3, pp. 51–80.

Alberola, E., J. Chevallier, and B. Chèze (2008). "Price drivers and structural breaks in European carbon prices 2005-2007". In: *Energy Policy* 36.2, pp. 787–797.

Benz, E. and S. Trück (2009). "Modelling the price dynamics of CO2 emission allowances". In: *Energy Economics* 31.1, pp. 4–15.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity". In: *Journal of Econometrics* 31.3, pp. 307–327.

— (1987). "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return". In: *The Review of Economics and Statistics* 69.3, pp. 542–547.

Byun, S. J. and H. Cho (2013). "Forecasting carbon futures volatility using GARCH models with energy volatilities". In: *Energy Economics* 40, pp. 207–221.

Capoor, K. and P. Ambrosi (2006). *State and Trends of the Carbon Market 2006*. Washington DC: World Bank.

Chevallier, J. (2009). "Carbon Futures and Macroeconomic Risk Factors: A View from the EU ETS". In: *Energy Economics* 31.4, pp. 614–625.

— (2010). "EUAs and CERs: Vector Autoregression, Impulse Response Function and Cointegration Analysis". In: *Economics Bulletin* 30.1, pp. 558–576.

— (2011). "A model of carbon price interactions with macroeconomic and energy dynamics". In: *Energy Economics* 33.6, pp. 1295–1312.

Christiansen, A.C., A. Arvanitakis, K. Tangen, and H. Hasselknippe (2005). "Price determinants in the EU emissions trading scheme". In: *Climate Policy* 5.1, pp. 15–30.

Daskalakis, G., R.N. Markellos, and D. Psychoyios (2009). "Modeling CO2 emission allowance prices and derivatives: Evidence from the European trading scheme". In: *Journal of Banking & Finance* 33.7, pp. 1230–1241.

Ellerman, A. D. and B. K. Buchner (2007). "The European Union Emissions Trading Scheme: Origins, Allocation, and Early Results". In: *Review of Environmental Economics and Policy* 1.1, pp. 66–87.

Ellerman, A. D. and P. L. Joskow (2008). *The European Union's Emissions Trading System in perspective*. Pew Center on Global Climate Change.

Engle, R. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* 50.4, pp. 987–1007.

European Commission (2013). *EU ETS Fact Sheet*. Accessed 4 January 2014. URL: http://ec.europa.eu/clima/publications/docs/factsheet_ets_en.pdf.

Fama, E.F. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work". In: *The Journal of Finance* 25.2, pp. 383–417.

Glosten, L. R., R. Jagannathan, and D. E. Runke (1993). "On the relation between the expected value and the volatility of the nominal excess return on stocks". In: *The Journal of Finance* 48.5, pp. 1779–1801.

Hall, A. (1994). "Testing for a unit root in time series with pretest data-based model selection". In: *Journal of Business & Economic Statistics* 12.4, pp. 461–470.

IntercontinentalExchange Group (2011). *ICE futures Europe announces daily volume record for ECX EUA futures*. [Press release]. Accessed 7 January 2014. URL: http://ir.theice.com/investors-and-media/press/press-releases/press-release-details/2011/ICE-Futures-Europe-Announces-Daily-Volume-Record-for-ECX-EUA-Futures/default.aspx.

Kossoy, A. and P. Guigon (2012). *State and Trends of the Carbon Market 2012*. Washington DC: World Bank.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" In: *Journal of Econometrics* 54.1, pp. 159–178.

Mansanet-Bataller, M., A. Pardo, and E. Valor (2007). "CO_2 Prices, Energy and Weather". In: *The Energy Journal* 28.3, pp. 73–92.

Mansanet-Bataller, M., J. Chevallier, M. Hervé-Mignucci, and E. Alberola (2011). "EUA and sCER phase II price drivers: Unveiling the reasons for the existence of the EUA-sCER spread". In: *Energy Policy* 39.3, pp. 1056–1069.

Marquardt, D.W. (1970). "Generalized inverses, ridge regressions, biased linear estimation, and nonlinear estimation". In: *Technometrics* 12.3, pp. 591–612.

Paolella, M.S. and L. Taschini (2008). "An econometric analysis of emission allowance prices". In: *Journal of Banking & Finance* 32.10, pp. 2022–2032.

Rickels, W., D. Görlich, and G. Oberst (2010). "Explaining European Emission Allowance Price Dynamics: Evidence from Phase II". In: *Kiel Working Papers* 1650.

Schwarz, G.E. (1978). "Estimating the dimension of a model". In: *Annals of Statistics* 6.2, pp. 461–464.

Schwert, G.W. (1989). "Tests for Unit Roots: A Monte Carlo Investigation". In: *Journal of Business & Economic Statistics* 7.2.

Trück, S., W. Härdle, and R. Weron (2012). "The relationship between spot and futures CO_2 emission allowance prices in the EU-ETS". In: *Midwest Finance Association 2013 Annual Meeting Paper*.

Tsay, R.S. (2010). *Analysis of Financial Time Series*. 2nd ed. Hoboken, New Jersey: John Wiley & Sons.

Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. Uppsala: Almquist & Wiksell.

A Methods

A.1 ARMA and GARCH models

We use several methods for estimating returns. Beginning with the simplest autoregressive models, $AR(r)$ where r denotes the amount of lag steps considered. The formula for the $AR(r)$ model is

$$y_t = c + \sum_{i=1}^{r} \phi_i y_{t-i} + \epsilon_t \tag{4}$$

where c and ϕ_i are real constants, typically estimated using an ordinary least squares (OLS) procedure, and y_t is the return at time t. The residuals ϵ_t are assumed to be independent and identically distributed with zero-mean and constant variance. An extension of this is an autoregressive moving average (ARMA) model, proposed by Whittle (1951), which also depends on lagged residuals of the time series. The formula for an $ARMA(r, m)$ model is

$$y_t = c + \sum_{i=1}^{r} \phi_i y_{t-i} + \sum_{j=1}^{m} \theta_j \epsilon_{t-j} + \epsilon_t \tag{5}$$

where θ_j are constant parameters. However, the assumption that variance is constant over time (i.e. the time series is homoskedastic) is not always true. As shown in for example Benz and Trück (2009), there is reason to believe that this is not the case for emission allowance returns. The autoregressive conditional heteroskedastic $(ARCH(q))$ model proposed by Engle (1982) deals with this problem. In the simplest case it is assumed that the time series y_t is zero-mean, and then one models the residuals ϵ_t split up into a stochastic part, u_t, and a time-dependent standard deviation term σ_t. The $ARCH(q)$ model considers q lagged terms

$$y_t = \epsilon_t, \text{ and } \epsilon_t = u_t \sigma_t, \text{ with } \sigma_t^2 = a_0 + \sum_{k=1}^{q} a_k \epsilon_{t-k}^2 \tag{6}$$

where a_k are the parameters of the model. It is assumed that u_t are i.i.d. with zero mean and variance equal to 1. Typically, the distribution of u_t is assumed to be either standard normal or standardized t. By using not only the past ϵ_t-values but also a moving average of past σ_t to model the conditional variance, a generalized autoregressive conditional heteroskedastic model $(GARCH(p, q))$ is obtained. This model, put forward by Bollerslev (1986) is defined as

$$y_t = \epsilon_t, \text{ and } \epsilon_t = u_t \sigma_t, \text{ with } \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{7}$$

where ω, α_i and β_j are real constants and ϵ_t are as before. The parameters have to satisfy $\Sigma\alpha_i + \Sigma\beta_j < 1$ to ensure that the variance stays finite, and $\alpha_i, \beta_j \geq 0$, $\omega > 0$ so that we have stationarity and positive variance. The GARCH model is quite sophisticated regarding variance, but the actual values of y_t are just assumed to be i.i.d., zero-mean random variables. By combining an ARMA(r, m) process for the mean with a GARCH(p, q) model for the variance one obtains an auspicious model for the price dynamics.

A.2 MSE and MAE formulas

The formula for calculating the mean squared error is

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2$$

where $\hat{Y}_i$ is a vector of predicted values, Y_i a vector of the actual values and n is the number of predictions. To calculate the mean absoulte error we use

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{Y}_i - Y_i|$$

B   Figures
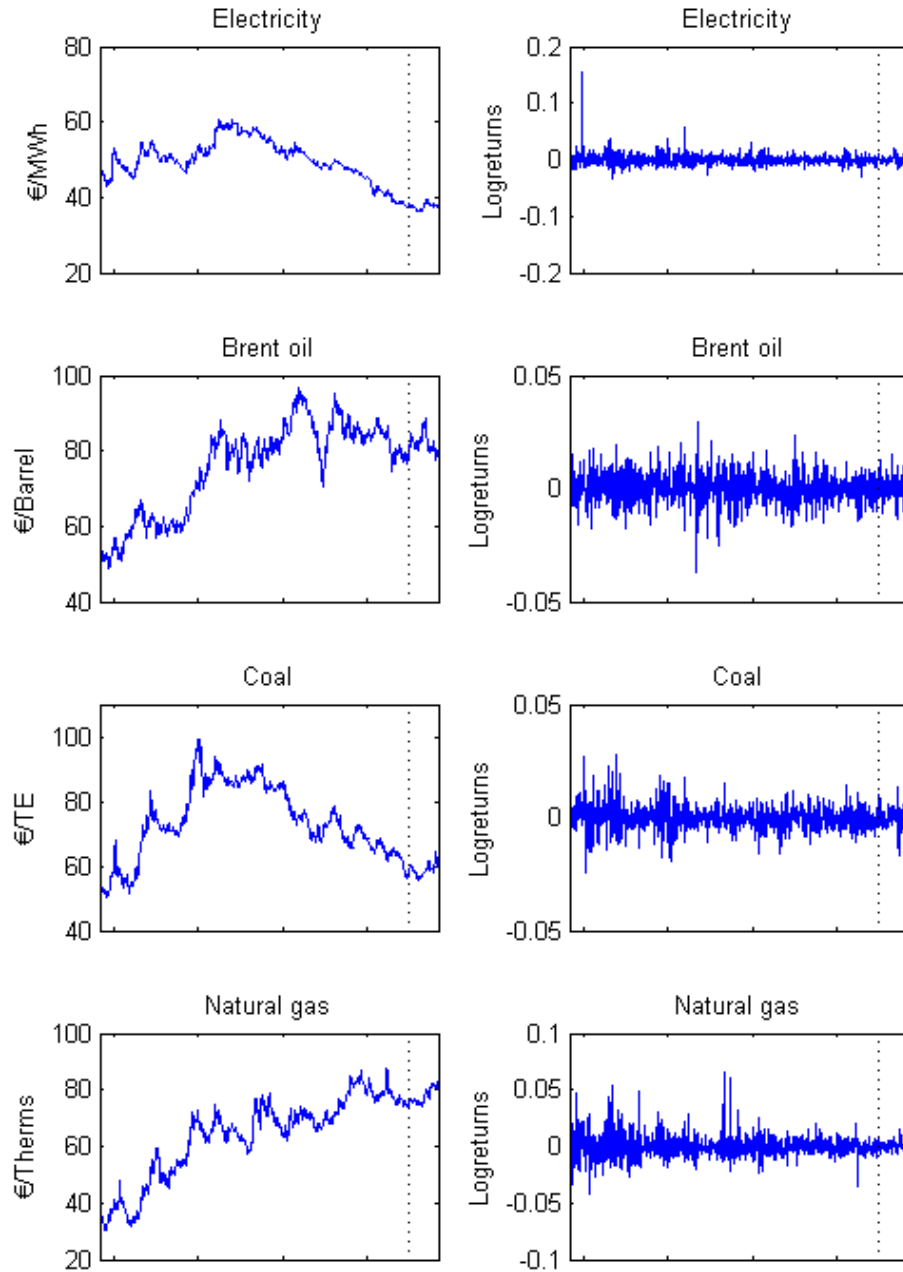
Figure 6: The price and returns of our exogenous variables, i.e. electricity, Brent oil, coal and natural gas prices converted to EUR, from November 2009 to November 2013. The dashed line indicates where the out-of-sample period starts.

C Tables

C.1 Variables and abbreviations

Abbreviation	Explanation
ADF	Augmented Dickey-Fuller
AIC	Akaike information criterion
AR	Autoregressive
ARA	Amsterdam-Rotterdam-Antwerp
ARCH	Autoregressive Conditional Heteroskedasticity
ARMA	Autoregressive Moving Average
ARMAX	ARMA with exogenous inputs
BIC	Bayesian information criterion
CO_2	Carbon dioxide
ECB	European Central Bank
ECX	European Climate Exchange
EEX	European Energy Exchange
ETS	Emissions Trading System
EU	European Union
EUA	European Union Allowance
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GARCHX	GARCH with exogenous inputs
GHG	Greenhouse gas
GJR-GARCH	Glosten-Jagannathan-Runkle GARCH
ICE	IntercontinentalExchange
IID	Independent identically distributed
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
LLF	Log-likelihood function
LM	Lagrange multiplier
LR	Likelihood ratio
MA	Moving Average
MAE	Mean absolute error
MSE	Mean squared error
OLS	Ordinary least squares
TGARCH	Threshold GARCH
UK	United Kingdom
US	United States
VIF	Variance inflation factor

Table 9: Summary of the abbreviations used in this thesis, ordered alphabetically.

Variable	Explanation
r	The number of autoregressive (AR) terms in the mean equation of the ARMAX-GARCH model.
m	The number of moving average (MA) terms in the mean equation of the ARMAX-GARCH model.
b	The number of exogenous inputs in the mean equation of the ARMAX-GARCH model.
p	The number of lagged variances (GARCH terms) considered in the variance equation of the ARMAX-GARCH model.
q	The number of lagged squared residuals (ARCH terms) considered in the variance equation of the ARMAX-GARCH model.
y_t	Logreturn of the EUA continuous futures contract at time t. We have that $y_t = \ln(s_t) - \ln(s_{t-1})$.
c	Constant in the mean equation of the ARMAX-GARCH model.
ϕ_i	The AR parameter at lag i.
θ_j	The MA parameter at lag j.
η_k	The exogenous variable coefficient at lag k.
$d_{k,t-1}$	Logreturn of the k:th exogenous variable at time $t-1$.
ϵ_t	The residual of the EUA return at time t. The difference between the actual and the estimated value of y_t.
u_t	Independent and identically distributed random variables that are zero-mean and have unit variance.
σ_t	The standard deviation of the EUA return at time t.
α_0	Constant in the variance equation of the ARMAX-GARCH model.
α_i	The ARCH parameter at lag i.
ϕ_j	The GARCH parameter at lag j.
$\hat{L}$	Maximized value of the log-likelihood function.
k	Number of parameters in a model.
n	Amount of observations.
D	Likelihood ratio test statistic.
Q	Ljung-Box Q-test statistic.
$\hat{\rho}_k$	Sample autocorrelation at lag k.
l	Number of lags considered in the Ljung-Box Q-test.
s_t	Price of an EUA continuous futures contract at time t.
ν	Degrees of freedom for the t-distribution.

Table 10: Summary of the variables used in this thesis, in order of appearance.

C.2 Summary statistics for exogenous variables

Dataset	Size	Mean	Min	Max	Std. Dev.	Skew.	Kurt.
Electricity							
In-sample	953	-0.0003	-0.0326	0.1535	0.0098	4.5703	68.1653
Out-of-sample	90	$-3.6 \cdot 10^{-5}$	-0.0166	0.0199	0.0067	0.3922	3.4114
Brent oil							
In-sample	953	0.0002	-0.0369	0.0293	0.0067	-0.1600	4.7084
Out-of-sample	90	$-2.2 \cdot 10^{-5}$	-0.0135	0.0148	0.0055	-0.0677	3.4085
Coal							
In-sample	953	$5.4 \cdot 10^{-5}$	-0.0242	0.0277	0.0056	0.3312	5.9448
Out-of-sample	90	0.0002	-0.0166	0.0139	0.0046	-0.3876	5.0420
Natural gas							
In-sample	953	0.0003	-0.0416	0.0661	0.0097	1.1093	9.7272
Out-of-sample	90	0.0004	-0.0072	0.0156	0.0037	1.2932	6.8613
Paper							
In-sample	953	0.0001	-0.0118	0.0233	0.0016	4.4210	65.8129
Out-of-sample	90	-0.0001	-0.0052	0.0033	0.0009	-1.7555	15.4867
Stock index							
In-sample	953	0.0002	-0.0488	0.0691	0.0113	-0.1460	6.1663
Out-of-sample	90	0.0012	-0.0181	0.0231	0.0066	0.3344	4.2927

Table 11: Summary statistics for the exogenous variables returns. *Std. Dev.* refers to standard deviation, *Skew.* to the skewness, and *Kurt.* to the kurtosis.

C.3 Stationarity

ADF	t-statistic	p-value	lags
EUA	-9.4781	<0.01	15
Electricity	-15.4545	<0.01	3
Oil	-23.1724	<0.01	1
Gas	-18.8322	<0.01	3
Coal	-16.0751	<0.01	3
KPSS			
EUA	0.0915	>0.1	7
Electricity	0.2953	>0.1	7
Oil	0.1581	>0.1	7
Gas	0.0761	>0.1	7
Coal	0.4424	0.0588	7

Table 12: Statistics for the Augmented Dickey-Fuller (ADF) test and the KPSS test on the logarithmic return series. The low p-values of the ADF test means that we can reject the null hypothesis of a unit root at the 1% significance level. The KPSS test confirms stationarity at the 5% significance level.

C.4 ARCH LM test

	F-statistic	p-value
AR(0)	31.210	0.00183
AR(1)	31.234	0.00182
AR(2)	27.502	0.00654
AR(3)	27.876	0.00577
AR(4)	27.246	0.00712
MA(1)	31.240	0.00181
MA(2)	28.633	0.00447
ARMA(1,1)	32.696	0.00107
ARMA(2,2)	28.203	0.00517

Table 13: Results from the ARCH LM test at 12 lags.

C.5 Model evaluation with errors normally distributed

Model	k	$\hat{L}$	BIC
AR(0)	2	1802.7	-3591.7
AR(1)	3	1803.2	-3585.8
AR(2)	4	1813.9	-3600.3
AR(3)	5	1815.2	-3596.1
AR(4)	6	1819.5	-3597.9
MA(1)	3	1803.4	-3586.2
MA(2)	4	1811.7	-3597.2
ARMA(1,1)	4	1805.2	-3582.9
ARMA(2,2)	6	1819.3	-3597.5
GARCH(1,1)	4	2094.0	-4160.5
AR(1)-GARCH(1,1)	5	2094.8	-4155.2
MA(1)-GARCH(1,1)	5	2094.8	-4155.2
AR(2)-GARCH(1,1)	6	2094.8	-4148.4
MA(2)-GARCH(1,1)	6	2094.8	-4148.4
ARMA(1,1)-GARCH(1,1)	6	2094.8	-4148.4
ARMA(2,2)-GARCH(1,1)	8	2095.5	-4139.4
GARCH(2,2)	6	2101.1	-4161.0
GJR-GARCH(1,1)	5	2097.1	-4160.0

Table 14: Basic performance evaluation of the estimated models, with errors normally distributed. k is the number of parameters, $\hat{L}$ is the maximized value of the log-likelihood objective function, and BIC is the Bayesian information criterion. When comparing these values to the those in table 3, where we assume that the errors are t-distributed, one sees that indeed the models with errors t-distributed give better $\hat{L}$ and BIC values.

C.6 Choosing exogenous variables

Variable	η_1	S.D.	p-value
Electricity	-0.2022	0.0610	0.0010
Oil	-0.2479	0.0859	0.0040
Coal	-0.4037	0.1008	0.0001
Gas	-0.2696	0.0569	0.0000
Paper	0.3017	0.4150	0.4673
Stock index	-0.1056	0.0584	0.0708

Table 15: Results from ARMAX(0,0,1)-GARCH(1,1) evaluations. η_1 is the parameter for the exogenous variable and $S.D.$ is the standard deviation of the parameter. The variables with a p-value higher than 0.05 are disregarded.

C.7 Correlation of energy variables

	Electricity	Oil	Coal	Gas
Electricity	1			
Oil	0.1457	1		
Coal	0.3720	0.0466	1	
Gas	0.4121	0.1245	0.2882	1

Table 16: Cross-correlations of exogenous variables.

C.8 Variance inflation factors

Variable	R^2	VIF
Electricity	0.248	1.330
Oil	0.027	1.027
Coal	0.195	1.243
Gas	0.161	1.191

Table 17: Variance inflation factors for our four exogenous variables. If the VIF value is higher than 5 then problematic multicollinearities may be present.

C.9 ARMAX-GARCH parameter values

Model	c	η_e	η_o	η_c	η_g	α_0	α_1	β_1	ν
GARCH(1,1)	$-2 \cdot 10^{-4}$					$6 \cdot 10^{-6}$**	0.117***	0.883***	5.406***
	(0.001)					$(3 \cdot 10^{-6})$	(0.022)	(0.019)	(0.800)
Model 1	$-1 \cdot 10^{-4}$	-0.202***				$6 \cdot 10^{-6}$**	0.117***	0.883***	5.304***
	(0.001)	(0.061)				$(3 \cdot 10^{-6})$	(0.022)	(0.019)	(0.794)
Model 2	$-1 \cdot 10^{-4}$		-0.248***			$6 \cdot 10^{-6}$**	0.113***	0.887***	5.348***
	(0.001)		(0.086)			$(3 \cdot 10^{-6})$	(0.021)	(0.019)	(0.787)
Model 3	$-1 \cdot 10^{-4}$			-0.404***		$5 \cdot 10^{-6}$**	0.115***	0.885***	5.414***
	(0.001)			(0.101)		$(3 \cdot 10^{-6})$	(0.021)	(0.019)	(0.801)
Model 4	$-4 \cdot 10^{-5}$				-0.270***	$5 \cdot 10^{-6}$**	0.109***	0.891***	5.203***
	(0.001)				(0.057)	$(3 \cdot 10^{-6})$	(0.021)	(0.018)	(0.756)
Model 5	$-1 \cdot 10^{-4}$	-0.175***	-0.215**			$6 \cdot 10^{-6}$**	0.112***	0.888***	5.265***
	(0.001)	(0.062)	(0.085)			$(3 \cdot 10^{-6})$	(0.021)	(0.018)	(0.782)
Model 6	$-1 \cdot 10^{-4}$	-0.105		-0.323***		$5 \cdot 10^{-6}$**	0.116***	0.884***	5.376***
	(0.001)	(0.071)		(0.110)		$(3 \cdot 10^{-6})$	(0.022)	(0.019)	(0.801)
Model 7	$-3 \cdot 10^{-5}$	-0.090			-0.227***	$5 \cdot 10^{-6}$**	0.110***	0.890***	5.206***
	(0.001)	(0.068)			(0.062)	$(3 \cdot 10^{-6})$	(0.021)	(0.018)	(0.764)
Model 8	$-5 \cdot 10^{-5}$		-0.218**	-0.378***		$5 \cdot 10^{-6}$**	0.110***	0.890***	5.321***
	(0.001)		(0.086)	(0.100)		$(3 \cdot 10^{-6})$	(0.021)	(0.018)	(0.781)
Model 9	$5 \cdot 10^{-5}$		-0.228***		-0.259***	$5 \cdot 10^{-6}$**	0.106***	0.894***	5.174***
	(0.001)		(0.084)		(0.057)	$(3 \cdot 10^{-6})$	(0.020)	(0.018)	(0.749)
Model 10	$-3 \cdot 10^{-5}$			-0.287***	-0.210***	$5 \cdot 10^{-6}$**	0.109***	0.891***	5.248***
	(0.001)			(0.105)	(0.061)	$(2 \cdot 10^{-6})$	(0.021)	(0.018)	(0.765)
Model 11	$-3 \cdot 10^{-5}$	-0.084	-0.207**	-0.316***		$5 \cdot 10^{-6}$**	0.111***	0.889***	5.298***
	(0.001)	(0.073)	(0.085)	(0.110)		$(3 \cdot 10^{-6})$	(0.021)	(0.018)	(0.782)
Model 12	$5 \cdot 10^{-5}$	-0.064	-0.219***		-0.229***	$5 \cdot 10^{-6}$**	0.106***	0.894***	5.172***
	(0.001)	(0.070)	(0.085)		(0.062)	$(3 \cdot 10^{-6})$	(0.020)	(0.018)	(0.754)
Model 13	$-3 \cdot 10^{-5}$	-0.022		-0.274**	-0.202***	$5 \cdot 10^{-6}$**	0.110***	0.891***	5.250***
	(0.001)	(0.076)		(0.111)	(0.064)	$(2 \cdot 10^{-6})$	(0.021)	(0.018)	(0.767)
Model 14	$6 \cdot 10^{-5}$		-0.212**	-0.266**	-0.206***	$5 \cdot 10^{-6}$**	0.105***	0.895***	5.187***
	(0.001)		(0.085)	(0.104)	(0.061)	$(2 \cdot 10^{-6})$	(0.021)	(0.018)	(0.767)
Model 15	$6 \cdot 10^{-5}$	-0.002	-0.212**	-0.265**	-0.205***	$5 \cdot 10^{-6}$**	0.105***	0.895***	5.188***
	(0.001)	(0.079)	(0.086)	(0.111)	(0.063)	$(2 \cdot 10^{-6})$	(0.020)	(0.018)	(0.752)

Table 18: Parameter values of ARMAX-GARCH models. η_e, η_o, η_c and η_g are coefficients for electricity, Brent oil, coal and gas respectively. Standard deviations are in parentheses. *** indicates significance at 1% level, ** at 5% and * at 10%.

C.10   Ljung-Box Q-test

Model	$Q(7)$		$Q(12)$		$Q(17)$	
	Q-statistic	p-value	Q-statistic	p-value	Q-statistic	p-value
GARCH(1,1)	3.4831	0.8370	5.8857	0.9217	6.9048	0.9847
AR(1)-GARCH(1,1)	3.5972	0.8248	6.0597	0.9130	7.0508	0.9828
AR(2)-GARCH(1,1)	3.5722	0.8275	5.9351	0.9193	6.8325	0.9856
Model 1	3.7597	0.8070	6.3156	0.8993	7.2698	0.9797
Model 2	3.6218	0.8221	6.0932	0.9113	7.1713	0.9812
Model 3	3.4111	0.8445	6.2019	0.9055	7.1660	0.9812
Model 4	3.7597	0.8070	6.3156	0.8993	7.2698	0.9797
Model 5	3.8323	0.7988	6.3982	0.8946	7.3709	0.9782
Model 6	3.5407	0.8309	6.3582	0.8969	7.2959	0.9793
Model 7	3.4924	0.8360	6.2144	0.9048	7.1020	0.9821
Model 8	3.6313	0.8211	6.4824	0.8898	7.4203	0.9774
Model 9	3.4641	0.8390	6.1778	0.9068	7.1023	0.9821
Model 10	3.2670	0.8592	6.1387	0.9089	7.0394	0.9830
Model 11	3.7012	0.8134	6.5573	0.8854	7.4726	0.9765
Model 12	3.5962	0.8249	6.3329	0.8983	7.2398	0.9802
Model 13	3.3084	0.8550	6.1881	0.9063	7.0864	0.9824
Model 14	3.4782	0.8375	6.3959	0.8948	7.2810	0.9796
Model 15	3.4821	0.8371	6.4009	0.8945	7.2855	0.9795

Table 19: Results from the Ljung-Box Q-test on the squared standardized residuals.