An analysis of the fundamental price drivers of EU ETS carbon credits

Johan Obermayer

Abstract

This paper attempts to shed further light on price formation in the EU ETS carbon credits market. It explores relationships between credits and energy complex assets, including electrical power, coal, natural gas, and oil. Relationships are analyzed using various statistical tools and methods, and explored in terms of fundamental economic relationships, correlation, and cointegration. Furthermore, the applicability of certain statistical tools, specifically correlation and multivariate regression, are examined. The switching price, according to the fundamentals theory, is found to be a poor indicator for valuing EUAs. Further, pairwise correlations between carbon credits, specifically EUAs and energy complex assets, are mostly found to be very noisy and weak. Power is found to be the the only asset with significant correlation to EUAs. Weak correlations lead to weak multivariate regressions. Given these results, it is questionable whether correlation is a relevant tool for measuring relationships in the EU energy complex. To that end, cointegration is explored as a more relevant and robust measure. It is found that EUAs are cointegrated with natural gas and oil, but, surprisingly, not with power.

Price forecasters should observe that day-to-day EUA prices move in sync with electrical power. In the longer term, the EUA price is linked to oil, and, to a lesser extent, to natural gas.

Acknowledgements

I would like to thank the employees of Tricorona AB for their support and interest during the course of this thesis. In particular, I wish to thank Frank Larsgård for sharing his invaluable experience as an active market practitioner. Furthermore, I would like to thank my supervisor at the Royal Institute of Technology (KTH), Harald Lang, for several hours of good mathematical discussions. Last but not least, I wish to thank my family and friends for their encouragement, support, and patience during the entire course of this thesis.

Contents

Chapter 1

Introduction

Since the industrial revolution, anthropogenic activity has caused the atmospheric content of greenhouse gases to rise steadily, breaking new highs yearly. It is widely believed that a 4 °C average temperature increase in the Earth's climate above today's average would have catastrophic effects and may lead to shifts in steady environmental patterns. There is much debate as to whether higher atmospheric greenhouse gas concentrations cause global warming, but few doubt that anthropogenic activity is causing the world around us to change.

General consensus led to the 1997 ratification of the Kyoto protocol by 174 countries, amongst which were all the EU countries (notably, however, the USA did not ratify the treaty). The protocol requires participating countries to reduce their yearly CO_2 (or CO_2 equivalent) emissions to 5% below 1990 levels. The treaty is intended to achieve "stabilization of concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system."[1] As part of the agreement, a group of fifteen EU countries agreed to meet their reduction targets collectively through a so-called cap-and-trade scheme dubbed the European Union Emissions Trading Scheme (EU ETS).

With the new presidential administration in the USA having CO_2 reductions as a policy cornerstone, tentative steps towards a full-scale emissions trading scheme are being taken. Thus it appears that, with the USA finally on board, emissions trading is here to stay. Some sources estimate that the market size of emissions could be up to €3 trillion by 2020.[2] The EU ETS is currently the largest emissions trading scheme in the world and will set the standard in years to come, as well as be imitated as further emissions markets blossom worldwide.

Building a fundamental understanding of the functioning of the market is essential in planning large scale investment decisions that will be needed in order to meet emissions reduction targets. This paper is a step in the direction of getting a clearer view of what drives prices within the EU ETS. In the business of carbon credit trading the ultimate tool would be able to predict prices based on market conditions. Since accurate prediction in this (and indeed any) free market is fundamentally impossible, it is the author's opinion that the closest one can get is by having a deep understanding of market forces and thereby an ability to act appropriately on new information.

[1] UNFCCC website

[2] Gardner (2008)

The aim of this paper is to provide insight into the structure of market forces active in the EU market for CO_2 emissions, and to test which statistical methodologies are meaningful and applicable to the energy complex price processes presented here. This is done through empirical analysis of the price processes of various CO_2 emissions contracts, and of their connection to the price processes of other commodities in the energy complex through fundamental economic relationships. As the market has only been up-and-running since February 2005, still little is known about the price setting forces that are active in the market. Whereas the behavior of individual market players can in certain cases be analyzed for patterns, the market in aggregate is much harder to analyze. The focus is on the German market where it is perceived that there is a high degree of fuel switching capacity.

Much effort has been put into keeping this paper as understandable and readable as possible. We will proceed in a stepwise manner and mostly build from basic principles. Chapter 2 presents the basics of how the EU ETS functions. Chapter 3 discusses the fundamentals theory that underlies the approach in this report. Chapter 4 briefly reviews previous studies on this subject. Chapter 5 thoroughly examines the relevant mathematics that will be used in the analysis. Before the analysis, the data that will be used is briefly presented and discussed in Chapter 6. Finally, in Chapter 7, the analysis is performed and conclusions are drawn. Chapter 8 summarizes and discusses the material covered in the thesis and suggests topics for further work in the area.

Chapter 2

A crash course in the workings of the EU ETS

The EU ETS covers over 11,000 industrial installations in 15 EU member states that together are responsible for 40% of the EU's greenhouse gas emissions. It is thus currently the largest emissions trading scheme in the world. The ETS is a highly complex market, and a basic understanding of how it works is needed before it is possible to proceed with the analysis.

2.1 Cap-and-trade

The EU ETS is a so called cap-and-trade scheme. A cap-and-trade scheme, in short, designates a set of central authorities that control the total yearly allowed emissions (the cap). The cap is then tightened year-on-year in order to meet reduction targets. Those companies that manage to reduce emissions below their individual cap are allowed to sell their unused emission allowances to other companies that don't meet their caps.

In the EU ETS, the central authority is the United Nations Framework Convention on Climate Change (UNFCCC).[1] The UNFCCC allocates certificates for units of CO_2 emissions allowances, or carbon credits, to polluting industries. These industries must at the end of each year return a number of credits corresponding to the total amount of verified emissions produced from operations during that year. Credits come in units of one ton of CO_2 equivalent (CO_2e) emissions. Thus, a company covered by the ETS that pollutes with 1000 tons of CO_2e emissions during one year must return 1000 carbon credits at the end of that year, regardless of how many credits the company was allocated at the beginning of that year.

A company that is allocated fewer credits than it expects to produce must either a) reduce production in order to reduce emissions, b) buy additional credits from the market, or c) acquire new technology which allows the company to pollute less per unit of production. The course of action that a company chooses depends on the economics of that company. A company that can reduce emissions cheaply may sell surplus credits thus gained, and in addition thereby turn a profit on te allocated credits. Credits can be freely traded, and market forces set prices.

[1] This is highly simplified. In fact, in addition to the UNFCCC, there are several regulatory institutions included at the EU and national levels. A good description can be found at `http://ec.europa.eu/environment/impel/pdf/good_practice.pdf`

Each year the UNFCCC reduces the total number of freely allocated credits. This eventually forces all industries covered by the scheme to reduce emissions. The hope of the scheme is that the emissions reduction should come through technological improvements rather than reduction of production. Only technological improvement will, as a whole, cause less pollution per unit of produced goods.

The alternative to a cap-and-trade scheme is a flat tax on pollution. In a flat-tax scheme, the central authority would be able to control the price of pollution, but have no influence over the amount that is actually polluted. Since in a cap-and-trade scheme the total emissions is controlled, market forces set the price of pollution, which, on the whole, leads to emission reduction occurring where it is the most economical. This in turn leads to emissions reduction at the lowest cost to society. The situation is illustrated in figure 2.1 by the marginal emissions abatement curves of two companies.

Company B faces a higher marginal cost of compliance compared with company A. Here the advantage of the cap-and-trade scheme is clearly visualized. The UNFCCC sets the cap which creates differing costs of compliance for the two market players. A is able to reduce emissions, creating an allowance surplus. For B it is cheaper to buy credits than to abate internally. A thus sells its surplus emissions allowances to B and turns a profit. In the long run, however, B has an incentive to reduce anyway since the overall cap will be tightened year-on-year leading to ever rising costs of compliance. In a flat tax scheme, the price of emissions would be fixed, but the actual emissions for each company would be dictated by the marginal abatement curve for each company.

The ETS significantly alters the competitive landscape in the EU. Opponents argue that EU industries are put at a competitive disadvantage to the same industries abroad and therefore outside the ETS. This could cause companies to either move production abroad or increase production in plants based abroad, and thereby never actually achieve emissions abatement. Time will tell if this is the case, but the EU has clearly chosen to take this risk, citing that the risk of climate change is much higher. Also, opponents argue that poorer countries within the EU are placed at a disadvantage, which can severely hamper economic development. While this naturally is the case, others argue that these countries are, as members of the EU, recipients of other economic aid which on the whole offset the higher production costs.

As a side note, the EU ETS is based on the experience gained from a previous cap-and-trade system, namely the NO_2 and SO_2 market developed during the 1990s in the USA as a way of tackling problems associated with acid rain. The system was highly successful, giving proponents of the EU ETS faith in the capacity for a cap-and-trade scheme to actually produce significant emissions abatement.

2.2 Projects - increasing the supply of carbon credits

An essential aspect of the EU ETS is the increase in supply of carbon credits through emissions reductions in developing countries not covered by the ETS. The idea is that reducing emissions in developing countries is cheaper, and since global warming is indeed a global problem, emissions reduction should be incentivized regardless of geographic location. For example, a hydropower dam constructed in China (thus outside the EU ETS and Europe) 'saves' the global atmosphere from CO_2 that would otherwise have been created from a polluting coal fired power plant, and is thus granted a corresponding number of carbon credits.

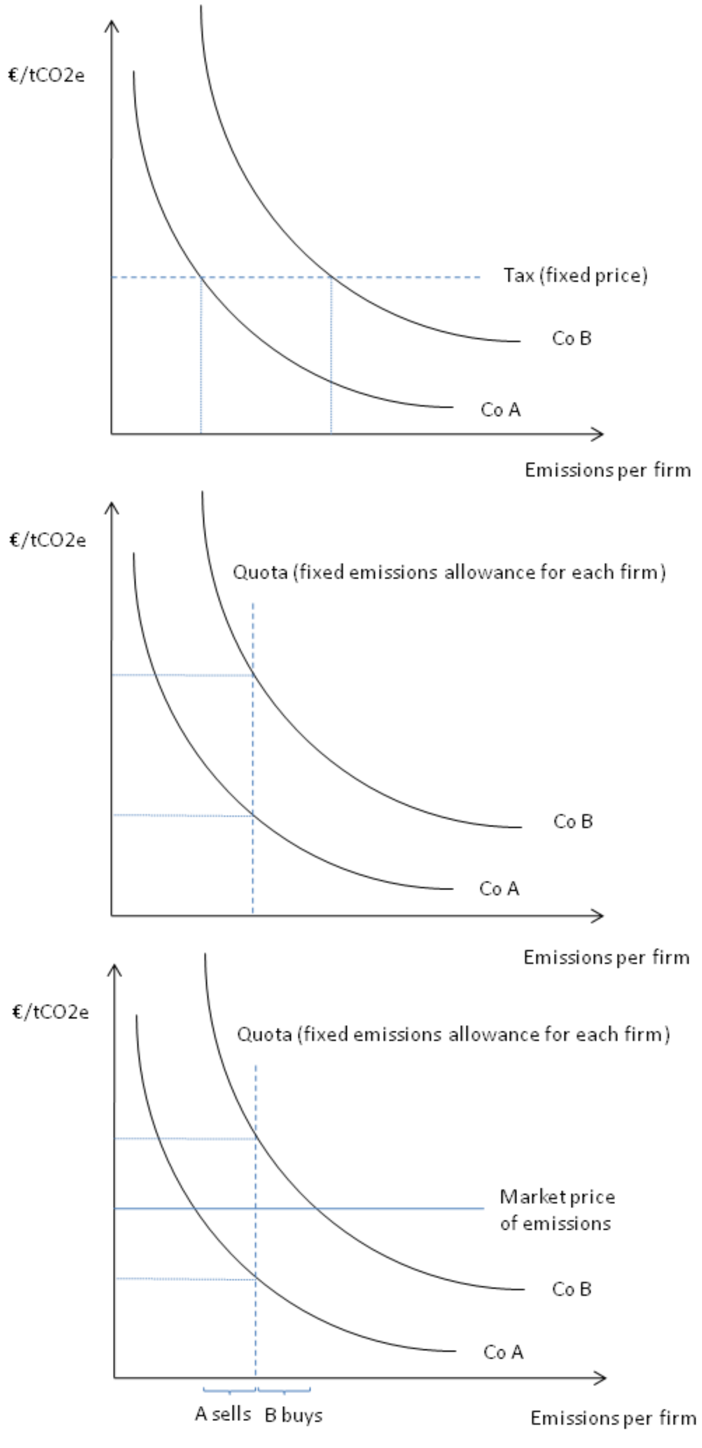

Figure 2.1: The marginal emissions abatement curves of two polluting companies

These credits which come from abatement in developing countries are called Certified Emission Reductions (CERs). CERs act as a supply increase on top of the free credits allocated to polluting industries by the UNFCCC. They may, however, only partly be used by industries in the ETS to offset emissions above the cap. The system of incorporating CERs into the ETS is called the Clean Development Mechanism (CDM) and is a part of the so called Kyoto Linking Directive. A simplified schematic of the above presentation is given in figure 2.2.

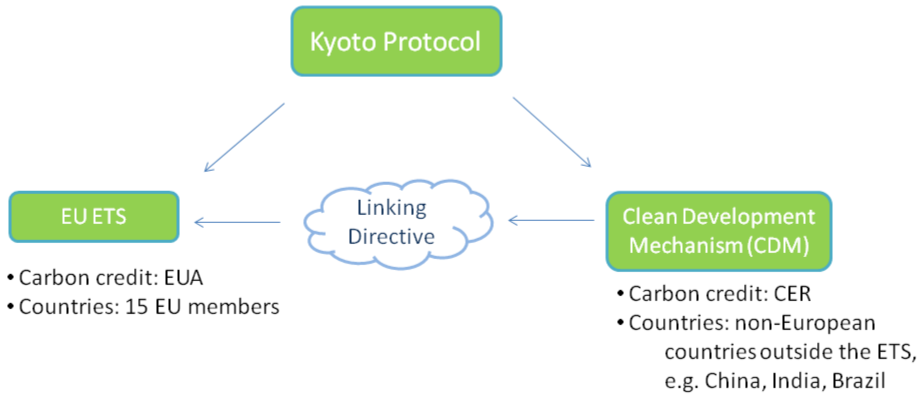

Figure 2.2: A simplified schematic of a part of the EU ETS

2.3 Market players

Market players can broadly be divided into three categories - energy sector, industrial sector, and speculators. The energy sector needs carbon credits to cover emissions that come from burning fossil fuels for producing power and heating. The industrial sector consumes power to produce goods. The industrial sector doesn't just implicitly pollute through power consumption, but can directly pollute as well. The energy and industrial sectors are thus 'compliance buyers', buying credits to cover emissions. Speculators of all varieties buy (or produce in the case of generating projects) in the hope of rising prices. The bullish view is that since the carbon credit quota is reduced year-on-year, prices should tend to rise in the medium term until technological development, leading to higher output per unit CO_2 produced, catches up.

2.4 The three phases

The establishment of the ETS is a dynamic process and has been set up in three phases. The regulatory framework is continually being revised as various systems are found to be effective or not, and future regulations are by no means set in stone. Since the market is an artificial one with no natural demand for carbon credits, the playing field depends entirely on the legal and regulatory framework. It is thus up to the UNFCCC and EU to provide stability and build faith in the system in order for the large scale private investments necessary to tackle climate change to take place.

Phase I from 2005-2007 may be considered the 'trial phase' in which administrative and regulatory bodies were put on-line. In phase I, all installations included in the scheme were given carbon credits corresponding to 100% of their respective emissions. Unfortunately, after a year of operation, it was found out that most installations had been oversupplied with carbon credits on the basis of falsely calculated operating 'baselines'. Naturally, the market then collapsed making credits worthless. However, since the credits from phase I couldn't be carried over, or 'banked', to phase II, the futures market for phase II credits still maintained high prices. This was because the market belief was that there would be adjustments made so as to not oversupply installations in phase II. Phase I is a prime example of policy failure leading to a market crash in this artificial environment, but also a vindication of the cap-and-trade system as providing economic transparency and thus a speedy signal to policy makers.

Phase II is the phase we are currently in, and runs from 2008-2012. Freely allocated credits are marginally fewer than forecasted operating baselines. Credits from this phase are bankable to phase III, meaning that no price fall is expected. However, the recent global financial crisis has caused industrial production to slow down, decreasing the demand for power. This in turn means that less emissions are produced and, given that the annual quota of free carbon credits is fixed, has caused carbon credit demand to collapse. An interesting development during this phase has been that carbon credits have been used as a form of financing. Many actors have chosen to sell EUAs on the spot market while compensating by buying long futures. This has led to a 'super contango' market, where futures are priced higher than they would have been using cost-of-carry calculations. This demonstrates that market forces can be far disconnected from economic equilibrium relationships, and that carbon credits must be seen as their own asset class with their own dynamics. Finally, traded volumes during this phase have been many times higher than the volume needed to even out short and long allocations, indicating a high degree of speculation.

Phase III is set to run from 2013-2020. It is widely believed that the demand for carbon credits will pick up as the economy recovers from the financial crisis. The preliminary plan is that by 2020 there will be a 21% reduction in allowances compared with 2005. If, however, international agreement (with non-EU countries) is reached on definite reductions, this number could increase to 30%. The idea is that if all dominant industrial actors are subject to ETS-like constraints, the relative competitiveness of EU industry will not be harmed. Another change set to come during this phase is that a large portion of what were previously free allowances is to be auctioned. Market forces will thus guarantee a more effective allocation of allowances. Those companies that previously made windfall profits by selling oversupplied credits will no longer be able to do so. Past 2020, the EU ETS is set to expire if no further agreement is reached following the next 'Kyoto' conference in December 2009 in Copenhagen.

Chapter 3

The fundamentals theory of the EU ETS price formation

The price of carbon credits is determined by supply and demand. Total supply is dictated by the UNFCCC allocation quota which is lowered every year (with a small amount of additional supply created by carbon credit generating projects), and demand is determined by industrial[1] and energy sector production. However, individual European governments, acting somewhat protectively towards the industry sector, tend to oversupply the industry sector and undersupply the energy sector with carbon credits. Furthermore, power and heating represent about 50% of all installations covered by the ETS. In fact, the power sector accounted for about 60% of emissions in phase I.[2]

The energy sector effectively becomes the most dominant demand side player, and thus sets the market dynamics. In order to further analyze market dynamics, it is crucial to ask the question "how does the energy sector fundamentally behave with respect to carbon credits?". In this section the economic relationships governing power production are shown. These concepts will be revisited in more detail in the analysis section.

3.1 Dark spread and spark spread

The energy sector produces electrical power[3] and heating. Heating can be disregarded for the following discussion since it is commonly a by product of power production. Power is generally produced by burning fossil fuel in order to heat water to steam that drives turbines to produce electricity. The economic equation of power production is thus (somewhat simplified)

$$Profit = Electricity\, price - Fuel\, price$$

All terms are per mega watt hour (MWh) of electricity. The two major types of fuel used in Europe today are coal and natural gas. When the fuel is coal, the above profit is called the 'dark spread', and when the fuel is gas, it is called 'spark spread'. The dark and spark spread must cover operating expenses for power production as a whole to be profitable.

[1] So far, a number of industrial sectors are included in the ETS. The EU plans, however, to include transportation - notably airlines - in phase III.

[2] Sikorski (2009)

[3] In this paper, 'power' denotes electrical power.

In order to refine this crude equation, we note that the fuel price per MWh will depend on the thermal efficiency of the power plant. If we let ρ represent the thermal efficiency of a power plant,

$$Dark\ spread = P_{electricity} - P_{coal} \cdot \frac{1}{\rho_{coal}}$$

$$Spark\ spread = P_{electricity} - P_{gas} \cdot \frac{1}{\rho_{gas}}$$

The values of ρ_{coal} and ρ_{gas} will in reality vary from plant to plant, but the values will fall between 30%-40% for coal and between 50%-60% for gas. It can generally be said that gas fired power plants have higher thermal efficiency but a higher price of fuel per MWh. A power producing company that has both a coal fired and a gas fired power plant in its portfolio will (disregarding operating costs) choose to run the plant that has better economics during the foreseeable future[4]. As mentioned previously, the introduction of the ETS has drastically altered the competitive landscape in the EU. For power production, the effect has been to alter the economic equation. This will be treated in the next section.

3.2 Clean dark spread and clean spark spread

The new equations governing power production are dubbed *clean dark spread* and *clean spark spread*, and are given by

$$Clean\ dark\ spread = P_{electricity} - [P_{coal} \cdot \frac{1}{\rho_{coal}} + P_{CO_2} \cdot E_{coal}]$$

$$Clean\ spark\ spread = P_{electricity} - [P_{gas} \cdot \frac{1}{\rho_{gas}} + P_{CO_2} \cdot E_{gas}]$$

Here E stands for emissions and is given in terms of tons of CO_2 emitted per MWh of electricity produced, and P_{CO_2} stands for the price of CO_2 per ton. Since coal is known to be twice as polluting per MWh of electricity produced as gas, gas fired power production begins to look much more attractive in a carbon constrained economy.

3.3 Switching price

Given a dark spread and a spark spread, it is possible to calculate the necessary implicit CO_2 price that would cause gas fired power production to be as profitable as coal fired power production. The formula for calculating the implicit switching price is given by

$$Switching\ price = \frac{\frac{P_{coal}}{\rho_{coal}} - \frac{P_{gas}}{\rho_{gas}}}{E_{gas} - E_{coal}}$$

The switching price is fundamental to carbon credit pricing in that the marginal cost of a credit should in theory follow the switching price. The argument goes as follows: with increasing carbon credit prices, more and more power producers will start to face negative clean dark spreads and will decide to switch (assuming that they have both types of plants in their portfolio, or have the capacity to switch). With a high degree of switching to cleaner gas, demand for carbon credits decreases, creating downward pressure that causes carbon prices to stabilize at or somewhat above the implicit switching price.

[4]Some modern fossil fuel power plants can switch between coal and gas, burning whichever gives the best economic profit.

A central assumption in the argument above is that power prices are stable. In reality, it appears that many power producers seek to maintain fixed clean dark and spark spreads and choose to pass on the cost of carbon credits to the consumers through a higher power price. Furthermore, there is debate as to whether rising carbon prices cause power prices to increase (i.e. passing on the price), or conversely whether rising power prices (i.e. rising power demand and increased total power output) cause carbon prices to increase. This is a chicken-and-egg argument which is ultimately decided by each individual power producer.

Chapter 4

Literature study

While there is a myriad of companies that produce carbon credit analyses, there have only been a limited number of research articles published.

Benz et al. (2007) investigated the the use of various GARCH models in an a-theoretical approach, in an attempt at capturing the price volatility dynamics of carbon credits. GARCH models were found to be effective in modeling short term behavior.

Kanamura (2009) examined whether carbon credits should be classified as commodities or securities. It was found that carbon credits in many ways do not behave like commodities since they did not exhibit mean-reversion and contained no seasonality component. Carbon credits are, however, called commodities in the common sense of the word.

Mansanet-Bataller et al. (2006) performed a correlation analysis similar to the one performed in this paper, but with a somewhat different approach. Correlations were found to be much higher than what is found in this report, and is a point that will be challenged here. In fact, the applicability of correlation in general as a measure of links in the EU ETS will be challenged.

Mansanet-Bataller et al. (2008) analyzed whether EUAs significantly contributed to diversification in the Markowitz portfolio framework. It was found that significant diversification effects could be achieved by including carbon credits in a portfolio.

POMAR (2007) was a large joint venture project between the University of Helsinki and the Helsinki University of Technology with the goal of predicting carbon credit prices, as well as how to effectively analyze risk. This study makes the point that switching price is a relevant variable, a further point that will be challenged in this report. Finally, there have been multiple papers on how to price options on carbon credits or various carbon credit spreads.

This paper explores various price processes for market structure and, as mentioned, challenges a few preconceptions about the market behavior. Furthermore, presented methodologies are tested for applicability and usefulness in extracting information about the market. With the rise of the asset class, many attempts have been made at price forecasting. A common approach has been to examine various sorts of correlations with other assets or indexes. However, very little work has been done using 'cointegration'. This is one of the methodologies that will be studied in this report, and constitutes a cornerstone of the analysis.

Chapter 5

The math

This section will treat the mathematics that will be used in the analysis. The less mathematically interested reader can skip the formulas but should nonetheless read the section as these tools will be the foundation of the coming analysis. The goal, somewhat vaguely stated, is to gain an understanding for market structures. To this end, several mathematical tools will be used. These are: correlation, multivariate regression, and cointegration.

Correlation is a look at to what degree sample data, in our case a price time series, moves synchronously. Multivariate regression investigates how a time series varies as a function of several other time series. Finally, cointegration is a way of establishing if there exists a long term link between price processes that is not exposed using univariate or multivariate regression. In order to fully understand how these are applied to financial data, it is important to first have an understanding for basic time series analysis, as well as how correlation on sample data works. This is the topic of the first section.

5.1 Basic time series analysis

Financial data has the price of asset as the starting point. From price data, the returns are calculated as the percent change in price from one time period to the next,

$$\Delta S_t \equiv \frac{S_t - S_{t-1}}{S_{t-1}}$$

A commonly used approximation for small returns is

$$\Delta S_t \approx \ln S_t - \ln S_{t-1},$$

which is exact for continuously compounded returns. Price and returns data are fundamentally different, and to understand why, we need to understand the concept of stationarity. The logarithm of most financial data can be described as autoregressive, meaning that the current price is a coefficient α times the previous period's price with a stochastic adjustment

$$\ln S_t = \alpha \ln S_{t-1} + \epsilon_t,$$

where ϵ is the adjustment term. More precisely, this process is called autoregressive, or integrated, of order one, and is written AR(1), or I(1) in shorthand. When ϵ for each time period is independent of other periods and identically distributed with distribution

$\epsilon_t \sim N(0, \sigma^2)$, it is dubbed white noise. This is often taken as an assumption in financial data analysis.

An AR(1) process is generally only stable for $|\alpha| \leq 1$. For $\alpha = 0$ the process is just white noise centered around zero. For $|\alpha| = 1$ the process is a random walk that does not exhibit 'mean reversion', meaning that it does not center around a long run mean. For $|\alpha|$ increasing from zero to one, the level of mean reversion decreases. A process is called stationary when it exhibits mean reversion, i.e. for $|\alpha| < 1$.

The logarithm of price data usually has $\alpha = 1$. Further, we see from the above formulas that returns data are of the form

$$\Delta S_t \approx \ln S_t - \ln S_{t-1} = \epsilon_t$$

Returns data are thus stationary - the ΔS_t center around a common point, namely zero.

The reason that the formulas above work for financial time series is the efficient market hypothesis, and comes from basic option pricing models. In an efficient market, asset prices reflect all information available to that point. Future movements are therefore the result of new information or news, which is unpredictable. Expressed differently, the best prediction of tomorrow's price is just today's price. The fundamental assumption in this theory is that asset prices (under the risk neutral probability measure) follow the geometric Brownian motion (GBM)

$$dS_t/S_t = rdt + \sigma dB_t$$

where r an σ are constants that define the drift and volatility of the asset prices, and B_t is a Wiener process. Increments dB_t are thus normally distributed $N(0, dt)$. Applying Itô's formula to this yields

$$d \ln S_t = (r - \sigma^2/2)dt + \sigma dB_t$$

Expressing this in discrete time yields

$$\ln S_t = c + \ln S_{t-1} + \epsilon_t,$$

meaning that asset returns are modeled as $\Delta S_t \approx \ln S_t - \ln S_{t-1} = c + \epsilon_t$, where $c = r - \sigma^2/2$ and $\epsilon \sim IID\ N(0, \sigma^2)$. With $c = 0$ we obtain the random walk process which is used to model log prices in efficient financial markets, and we recover our above formulas for asset returns. If $c > 0$ the price trends upwards, and vice versa. In less than efficient markets, log returns may still be modeled as above, but the increments ϵ may be autocorrelated and therefore not IID.

5.2 Correlation

Correlation is a measure of the degree to which two variables move in sync. Correlation lies at the heart of multivariate regressions, which will be treated in the next section. The correlation between the random stochastic variables X and Y is denoted by $Corr(X, Y)$, or often just $\rho_{X,Y}$, and is defined as follows

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{E[XY] - E[X]E[Y]}{\sqrt{E[X^2] - E[X]^2}\sqrt{E[Y^2] - E[Y]^2}}$$

When taken on a sample, the Pearson correlation coefficient is the most commonly used and is defined as

$$r_{x,y} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

If the data column vectors $\mathbf{x}$ and $\mathbf{y}$ are normed, i.e. the mean is subtracted and the data is divided by the sample standard deviation, the sample correlation can be calculated as

$$r_{x,y} = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

If the normed sample data is put in a $T \times k$ matrix $\mathbf{X}$, the correlation matrix $\mathbf{V}$ of several normed time series can be written compactly as

$$\mathbf{V} = \mathbf{X^T X / T}$$

where T denotes the sample size.

Correlation has a very nice geometric interpretation which we will use later. If we define $\vec{X}$ and $\vec{Y}$ as the 'vectors' of two stochastic variables with squared lengths by $Var(X)$ and $Var(Y)$, and their mutual projections given by $Cov(X, Y)$, then the correlation coefficient will be the cosine of the angle between them.

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = cos(\theta)$$

The smaller the angle, the higher the correlation. For normed sample data $\mathbf{x}$ and $\mathbf{y}$, the interpretation is that the closer the column vectors are to each other in terms of angle, the higher the correlation.

The square of the sample correlation coefficient, called the coefficient of determination and denoted R^2, is the fraction of the variance in Y that is explained by a linear fit of X to Y,

$$R^2 = r_{x,y}^2$$

This should be read as "R^2 percent of the variation of Y is explained by the variation of X".

Having established how to calculate the (Pearson) correlation of a sample, it is now time to turn to how to measure correlation for the types of financial time series that we are interested in - namely price processes. When calculating the correlation between two financial time series, we are asking to what degree the changes in price data co-vary. If we read the word 'changes' as 'returns', we immediately see that it is crucial to perform the above correlation calculations on the returns data´, and not the original price data. Let S^1 denote the price of asset 1 and S^2 denote the price of asset 2, and let them be defined as

$$\Delta S^1 = \ln S_t^1 - \ln S_{t-1}^1 = c_1 + \epsilon_t$$

$$\Delta S^2 = \ln S_t^2 - \ln S_{t-1}^2 = c_2 + e_t$$

The correlation between two asset price time series is thus

$$Corr(\Delta S^1, \Delta S^2) = Corr(\ln S_t^1 - \ln S_{t-1}^1, \ln S_t^2 - \ln S_{t-1}^2) = Corr(\epsilon_t, e_t)$$

From this we see that it the degree to which the stochastic increments ϵ and e move synchronously that defines the correlation. Calculating correlations on raw price data is a common error and can lead to the appearance of high correlation where there in truth is none - this phenomenon is called spurious regression. We will illustrate the point further in the following example.

Consider two discrete-time fictional assets S^1 and S^2 that are defined as discrete geometric Brownian motions (GBM) with correlation $\rho_{1,2}$. Simulating asset returns with low correlation, $\rho_{1,2} = 0.3$, but that have a common trend $c_1 = c_2 = c$, figure 5.1 is obtained.[1] Since the price processes have a common trend one might be fooled to think that the correlation should be high where there in fact is very little. Furthermore, if the correlation is computed on the raw price data, we get spurious correlation $Corr(S^1, S^2) = 0.98$, which is entirely false. When calculated on the returns data, we get the real correlation $Corr(\epsilon, e) = 0.29$. The discrepancy from the fixed $\rho_{1,2}$ is accounted for by statistical variation, and goes to zero as the sample size increases. Similarly, diverging time series (for example where the trends are of opposite sign) can still be highly correlated. This situation is illustrated in figure 5.2 where the correlation is set to $\rho_{1,2} = 0.7$.

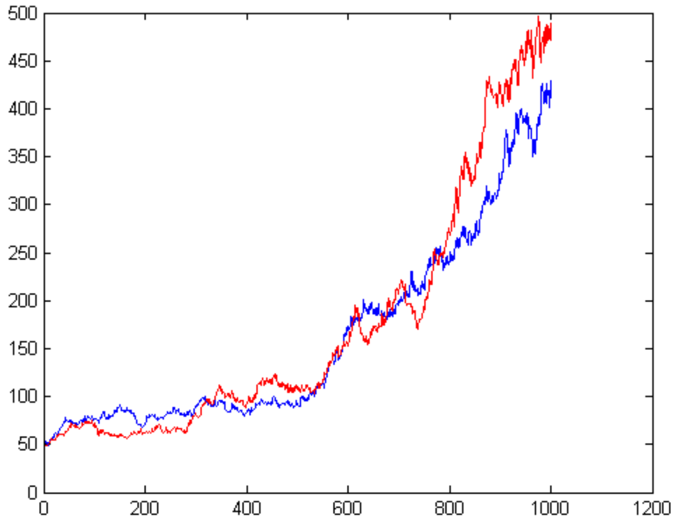

Figure 5.1: GBM time series simulation with low correlation

A final note on correlation is that the formula for calculating correlation gives equal weight to all observations. In dynamic financial data it would make more sense to place larger weight on recent data points. One common way of doing this is through *exponentially weighted moving average correlation* (EWMA correlation) which is defined as follows,

$$r_{x,y} = \frac{\sum \lambda^i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum \lambda^i (x_i - \bar{x})^2 \cdot \sum \lambda^i (y_i - \bar{y})^2}}$$

Here $\lambda \in (0, 1)$ is a factor which sets the rate of decay (of importance). High λ means that distant points are still weighed heavily, and vice versa. We will use $\lambda = 0.94$ in this paper.

[1] For a full discussion on how to simulate correlated asset prices see the appendix

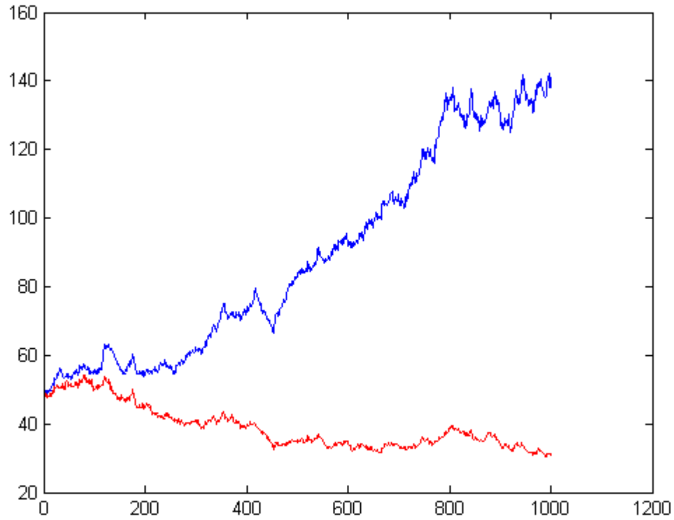

Figure 5.2: GBM time series simulation with high correlation

5.3 Linear regression

Linear regression is a test of whether, and how, a variable (called the dependent variable) is related to one or several other variables (called the independent variables). A linear regression is characterized by the way the dependent variable is seen as a function of the independent. Note that 'linear' does not refer to this straight line, but rather to the way in which the regression coefficients occur in the regression equation.

It is important to make the distinction between a linear relationship and a linear regression. In a linear relationship, the dependent variable does in fact vary linearly with the independent variables. In a linear regression, a linear model is fitted to sample data. Oftentimes the situation to which a linear regression is applied is not derived from a linear model. The power of a linear regression lies in its simplicity, and in some cases much can be said by applying a linear regression to a non-linear relationship. It is impossible to establish causality using linear regression, but knowing how two processes depend on each other is relevant none the less.

5.3.1 Definition

A linear relationship between $X_1, X_2, \ldots, X_k$ and Y is defined as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$$

where $\beta_1, \beta_2, \ldots, \beta_k$ are coefficients that measure the effect that the associated independent variable has on Y. Generally, a linear relationship isn't perfectly linear, so an error term ϵ is added:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \epsilon$$

In the multivariate case there are several independent variables, and in the univariate case there is only one (not counting the constant).

When sample data in a linear model is time series data, it is indexed with t. A whole sample data set can be written compactly using matrix notation. We let the vector $\mathbf{y}$ denote all samples of the variable being regressed $\mathbf{y} = (\mathbf{Y_1}, \mathbf{Y_2}, \ldots, \mathbf{Y_T})$, and row j of the matrix $\mathbf{X}$ corresponds to a single observation of all explanatory variables. The task thus becomes finding the β values that fit all samples Y_t and X_t in a linear model in the "best" possible way. Given that we have found the 'best' parameters $\hat{\beta}$, we can predict outcomes of Y based on new X's. The predicted Y_t's are denoted $\hat{Y}_t$, and are calculated as $\hat{Y}_t = \hat{\beta}X_t$. The difference between the true Y_t and the predicted $\hat{Y}_t$ is called the error (or residual) e_t, $e_t = Y_t - \hat{Y}_t$.

The best parameters would logically in some way minimize the error e_t, and the most common way this is done is through ordinary least squares (OLS). OLS estimators minimize the sum of the square error terms, $SSE = \sum e_t^2 = \mathbf{e^T e}$. One way of obtaining OLS estimators $\mathbf{b}$ (when using OLS the estimators $\hat{\beta}$ are denoted $\mathbf{b}$) is by solving a quadratic optimization problem, resulting in

$$\mathbf{b} = (\mathbf{X^T X})^{\mathbf{-1}} \mathbf{X^T y}$$

The result can also be obtained geometrically by noting that the parameter vector b that minimizes the euclidean norm of the vector of error terms, $\mathbf{e}$, must be orthogonal to the plane spanned by the column vectors in $\mathbf{X}$.

In the univariate case the result reduces to

$$b = \frac{\hat{\sigma}_{X,\mathbf{y}}}{\hat{\sigma}_X^2} = \frac{\rho \hat{\sigma}_y}{\hat{\sigma}_X}$$

where the hat denotes that the variable is an estimate based on sample data. A univariate linear regression is thus based almost entirely on the concept of correlation, and the multivariate case is an extension thereof.

5.3.2 Confidence level

Further calculations reveal that the estimated covariance matrix $V(\mathbf{b})$ for the components of $\mathbf{b}$ is calculated as

$$\hat{V}(\mathbf{b}) = \mathbf{s^2}(\mathbf{X^T X})^{\mathbf{-1}}, \; \mathbf{s^2} = \frac{\mathbf{e^T e}}{\mathbf{T - k}},$$

where T is the sample size (the number of samples of t) and k is the number of variables in the regression including the constant. This estimation can be used for calculating

confidence bounds on the components of $\mathbf{b}$,

$$\frac{b_i - \beta_i}{\sqrt{\sigma_{b_i}^2}} \sim t_\nu, \; \nu = T - k \tag{5.1}$$

$$\Rightarrow P\left(b_i - t_{\nu,\lambda}\sigma_{b_i} < \beta_i < b_i + t_{\nu,\lambda}\sigma_{b_i}\right), \; \lambda = (1-\alpha)/2$$

It is common to just want to know if a b_i value could statistically be zero at a set confidence level α (this is in fact a two sided hypothesis test with $H_0 : \beta = 0$ and $H_1 : \beta \neq 0$). To do this a t score can be calculated as

$$t^* = (b_i - 0)/\sqrt{\sigma_{b_i}^2}$$

If the t score is smaller than the critical values of the t distribution, i.e. $t_{\nu,\lambda}$, then the null hypothesis is rejected and vice versa.

Equation 5.1 is only valid under the assumptions that the explanatory variables are non-stochastic and the errors $\mathbf{e}$ are stationary, homoscedastic (constant variance), and not autocorrelated. The conditions can be lowered if the regressors X are uncorrelated with the errors e (since then the explanatory variables can be stochastic), and if the errors have the same distribution.

Note that one of the problems with linear regressions applied to financial data is that financial structures are prone to regime switching. A relationship that is found to be true during a certain period may be entirely different during another. This means that even if a multivariate linear relationship is found between a combination of independent variables and the dependent variable, it may be short-lived.

5.3.3 Goodness of fit

Using linear regression is possible to determine how much a set of chosen explanatory variables affects a chosen dependent variable. As mentioned previously, it is not necessary that the data are in fact based on an underlying linear model. After performing a regression, it is interesting to know how well a regression fits the observed data. In other words, how far does the regression go to explain variation in the observed $\mathbf{y}$'s, and how much of the variation must be absorbed by the residual errors $\mathbf{e}$? To that end there are a number of metrics. We will here consider three metrics: R^2, adjusted R^2, and Bayesian information criterion (BIC).

In linear regressions, as in correlation, it is possible to calculate the coefficient of determination R^2. In order to do so for a linear regression, it is necessary to understand the concept of analysis of variance (ANOVA). In an ANOVA analysis, there are three basic concepts, the total sum of squares (TSS), the explained sum of squares (ESS), and the residual sum of squares (RSS). These are mathematically given by $TSS = (\mathbf{y} - \bar{\mathbf{y}})^{\mathbf{T}}(\mathbf{y} - \bar{\mathbf{y}})$, $ESS = (\hat{\mathbf{y}} - \bar{\mathbf{y}})^{\mathbf{T}}(\hat{\mathbf{y}} - \bar{\mathbf{y}})$ and $RSS = e^T e$. Through calculations one can see that $TSS = ESS + RSS$. Now R^2 is simply the explained sum of squares divided by the total sum of squares:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Thus we can tell how much variation of $\mathbf{y}$ is explained by the variation of the whole regression bX

Adjusted R^2, R^2_{adj}, is a modification of regular R^2 that takes into account the number of explanatory variables included. Unlike R^2, R^2_{adj} increases only if the new term improves the model more than would be expected by chance. R^2_{adj} can be negative, and will always be less than or equal to regular R^2. It is defined as

$$R^2_{adj} = 1 - \frac{RSS/(T-k-2)}{TSS/(T-1)}$$

R^2_{adj} does not have the same interpretation as regular R^2, and is mainly useful for deciding if an explanatory variable improves the results to such an extent that it should be included in the regression.

The Bayesian information criterion, or BIC, is similar to adjusted R^2 in that it is used for selecting appropriate variables to use in a regression. The lower the BIC, the better the regression. For each new variable that is added, there is a penalty function that increases the BIC. Variables can thus be added and removed until the lowest BIC is found. For approximately normally distributed errors, BIC is defined as

$$BIC = T\ln(RSS/T) + k\ln(T)$$

5.4 Cointegration

Cointegration is a way of testing whether there exists a long term linear relationship between two processes. "It is unfortunate that many market practitioners still base their analysis of the relationships between markets on the very limited concept of correlation. Trying to model the complex interdependencies between financial assets with so restrictive a tool is like trying to surf the internet with an IBM AT." This is the opening remark of the chapter treating cointegration in Alexander (2001). As mentioned previously, correlation is the degree to which asset returns move synchronously. However, if there is a long term relationship between two assets, this will by definition be lost when differencing to obtain returns. Correlation will miss this relationship entirely. For this kind of dependency, cointegration is needed.

Whereas correlation treats asset returns, cointegration examines asset prices directly (although the logarithm of the prices is taken first). If two time series x and y are integrated of the same order, but a linear combination of them $g = x - \alpha y$ is stationary (i.e. integrated of order zero), then the two are cointegrated. Stationarity implies that the combination g exhibits mean reversion. Cointegration remedies the intuitive problem that co-trending series should be correlated - even if they have zero correlation they may be cointegrated. The analogy commonly used for explaining cointegration is of a man walking his dog with a bundgy-cord leash. At times the two will be closer, at times further away, and at times they may even cross, but generally they hold an average distance from each other.

In testing whether two time series are cointegrated, it must first be verified that both series are integrated of the same order. As we have seen, financial data are generally integrated of order one (I(1)), so in our case this will not present a problem. After this, α must be determined such that g is stationary, and in doing so g must be tested for stationarity. The most common way of finding α is by performing an *Engle-Granger* test.

The Engle-Granger test is a two step process. First an univariate OLS regression is performed on the log-price data, then the residuals are tested for stationarity. Note that it is not common to regress price data on each other since this leads to spurious regression. The only time this is valid is if the series are cointegrated. If the residuals e_t are stationary, α will result from the regression

$$y_t = c + \alpha x_t + e_t$$

How, then, is it determined if the residuals are stationary? One common test for stationarity in a time series is the Dickey-Fuller test (DF). The DF test is relatively straight forward. We know that financial data are generally AR(1) due to the efficient market hypothesis, and are thus not stationary ($\alpha = 1$ in $\ln S_t = c + \alpha \ln S_{t-1} + \epsilon_t$).

To test if $\alpha = 1$ for sample data we could perform a regular OLS regression of the lagged variable y_{t-1} on the non-lagged variable y_t and use a simple t-test to test if the resulting α is significantly different than one. Using this α directly, however, would cause severe bias in the case of a unit root. Instead we difference once, resulting in

$$\Delta y_t = c + (\alpha - 1)y_{t-1} + \epsilon_t$$

and, after performing a regression, test if the coefficient of y_{t-1} is significantly different than zero.

A refinement of the DF test is the *augmented Dickey-Fuller test* (ADF). In the ADF test, lagged variables are added in order to remove any autocorrelation in the errors, so that the OLS regression gives an unbiased estimate of the coefficient of y_{t-1}.[2] The mathematical formulation of the ADF is

$$\Delta y_t = c + \beta y_t + \alpha_1 \delta y_{t-1} + .. + \alpha_m \Delta y_{t-m} + \epsilon_t$$

with the hypothesis test

$$H_0 : \beta = 0,\ H_1 : \beta < 0$$

As before, the coefficient normed by its standard deviation is t_ν distributed with $\nu = T - k$ degrees of freedom. We will use two lags in performing the ADF in this paper, and a significance level of 95%.

There are many more advantages to using cointegration analysis rather than correlation analysis. For example, if data is sampled frequently in one series and infrequently in another, as is often the case with economic data, it is impossible to calculate correlation, but it is still possible to examine cointegration.

Once it is established that two variables are cointegrated, this information could be used practically in trading in what is called 'pairs trading'. The idea is that if the spread between two variables is shown to exhibit a long-run mean, the spread can be purchased when it is lower than average, and sold when it is higher than average. This is done by going long on the higher priced asset and shorting the lower priced asset when the spread is low, and vice versa when the spread is high.

Note, however, that just like for multivariate regressions, cointegrated relationships in financial data are prone to regime switching. This means that if a cointegrated relationship between two assets is established using historical prices, there is no guarantee that it will stay cointegrated in the future. Furthermore, even if two assets are cointegrated, the variables can drift apart for a long time before the spread reverts back to the long term mean.

5.5 Data problems - multicollinearity

In analyzing time series data, there are often many practical and theoretical problems, especially with data. One such problem which is prominent in this analysis is multicollinearity in explanatory variables when performing a linear regression. Multicollinearity means that sample data between two or more explanatory variables are significantly correlated. A rule of thumb is that if inter-correlations between explanatory variables are higher than the R^2 of the entire regression, multicollinearity will cause a problem.

[2]For a more thorough explanation see for example Alexander (2001)

The reason is that if two variables are highly correlated, the OLS regression has 'trouble' determining on which variable to place the coefficient weight that 'rightfully' belongs entirely to one of the variables. The formal description of this is that the OLS estimates of the coefficients lack robustness.

There are several ways of overcoming multicollinearity, and one of the most powerful is through principal component analysis (PCA). PCA has many uses and can be a highly effective tool for overcoming many data problems. PCA transforms correlated data series into uncorrelated data series, and orders them in terms of the total variability in the data. For a short mathematical description see the appendix. We will assume a certain familiarity with PCA in the following discussion.

The way we use PCA in overcoming multicollinearity is by transforming our original correlated data series into uncorrelated data series (the PCA algorithm), performing our multivariate linear regression on the uncorrelated data, and then converting the estimated OLS coefficients back in order to match our original data. The reason this works for correlated data is because, when performing the regression using uncorrelated data the OLS estimates are much more robust. To see this intuitively, we remind ourselves of the discussion above on the geometric interpretation of correlation as the angle between two sample data vectors. When the angle is small (i.e. highly correlated data), the plane that contains both vectors is 'unstable', resulting in non-robust OLS estimates. When the angle is large (i.e. for uncorrelated data), the plane becomes much more 'stable'.

In mathematical terms, we see that for the normed sample data X^*

$$\mathbf{X}^* = \mathbf{P}\mathbf{W}^{\mathbf{T}}$$

where P is the matrix of uncorrelated (i.e. orthogonal) principal components, and W is the matrix of factor weights. Next, a regular multivariate OLS regression is performed using the (uncorrelated) principal components:

$$\mathbf{y} = \mathbf{a} + \mathbf{P}\mathbf{b} + \mathbf{e}$$

The OLS estimate of the vector $\mathbf{b}$ is

$$\mathbf{b} = (\mathbf{P}^{\mathbf{T}}\mathbf{P})^{-\mathbf{1}}\mathbf{P}^{\mathbf{T}}\mathbf{y} = \boldsymbol{\Lambda}^{-\mathbf{1}}\mathbf{P}^{\mathbf{T}}\mathbf{y},$$

since $\mathbf{P}^{\mathbf{T}}\mathbf{P} = \boldsymbol{\Lambda}$ which is the diagonal eigenvalue matrix of $\mathbf{X}^{*\mathbf{T}}\mathbf{X}$. We will now transform this result into the corresponding regression coefficients for our original correlated data. Substituting $\mathbf{P} = \mathbf{X}^*\mathbf{W}$ into this, we obtain

$$\mathbf{y} = \mathbf{a} + \mathbf{X}^*\mathbf{W}\mathbf{b} + \mathbf{e}$$

If we now define

$$\Sigma = \begin{bmatrix} \frac{1}{\sigma_{P_1}} & & \\ & \frac{1}{\sigma_{P_2}} & \\ & & \ddots \end{bmatrix}$$

where σ_{P_i} denotes the sample standard deviation of principal component i, then we can write

$$\mathbf{y} = (\mathbf{a} - \mu^{\mathbf{T}}\boldsymbol{\Sigma}\mathbf{W}\mathbf{b}) + \mathbf{X}^*\boldsymbol{\Sigma}\mathbf{W}\mathbf{b} + \mathbf{e}$$

We have now recovered the regression coefficients that we are looking for, based on the results of the regression on the uncorrelated principal components.

5.6 A note on other analyses

A modern and powerful way of measuring dependence between random variables is through copulas. In our context, however, the market is much too young for any stable dependence structures to have been established, and a copula analysis would not reveal anything. Furthermore, if the options market were more mature and liquid it could be possible to calculate implicit correlations - i.e. the correlation that is perceived by market players. But this is again not the case.

Chapter 6

Data discussion

6.1 Datasets

Before proceeding with an analysis using the mathematical tools presented above, we must first choose relevant data to analyze. Choosing relevant data is far from trivial, especially in such a young market. The fundamentals theory dictates the basic market mechanism, the assets of interest are: carbon credit price, wholesale power price, coal price, natural gas price. These are essentially all commodities.

Commodities generally have the characteristic that they are difficult to store (as compared with corporate bonds, for example). They are therefore generally traded on futures markets, and not on spot markets. Gas, for example, is virtually impossible to store in large volumes and is delivered continuously through a large network of gas pipes directly from the source. In most commodities markets, purchasing (intended for delivery) and hedging is performed on a month ahead or year ahead basis. Spot markets (where they exist) are only used for final adjustments. The effect is that far ahead futures prices are relatively stable, whereas spot market data are extremely noisy and don't reflect market mechanisms.

It can generally be said that power companies purchase and hedge on a year ahead basis. Therefore, in this paper, front year futures contracts are used. We will examine the German energy market, and will therefore use price data relevant to Germany. Table 6.1 states the relevant commodities and their contracts. The data were obtained from SKM SYSPower (a Norwegian power and commodities market data provider) and Bloomberg.

Carbon credit	EUA
Power	European Energy Exchange (EEX) Peak Load
Coal	CIF API2 Index
Natural gas	Zeebrugge
Oil	Sullom Voe Brent Crude

Table 6.1: Relevant commodities contracts

EEX peak load power is used as it is perceived that fuel switching occurs on the marginal unit of power produced, and that fuel mix for base power is relatively constant. Other data that will be used are the DAX index and the seasonally adjusted industrial production index for the eurozone, as found under the ticker '*ECST*' on Bloomberg. These will be used as indicators for the state of the economy and aggregate industrial output, respectively.

6.2 Getting familiar with EUAs

The EUA carbon credit is the de facto trendsetter for EU carbon credits. EUAs are the most 'useful' credit type in terms of covering emissions. The correlation coefficient between this EUA data and the corresponding CER data for all periods is >0.9. For these reasons, and because CER futures data are only available since mid 2008, we choose to let the EUA credit be representative of all EU ETS credits. In order to get a feel for EUAs, we begin by plotting the historical Y+1 data in figure 6.1. (In this context Y+1 future refers to the future that is valid until the end of that year. During 2008, for example, the 2008 future, that started trading in 2005, will be used until it expires in Dec 2008, and the new 2009 contract takes over.)

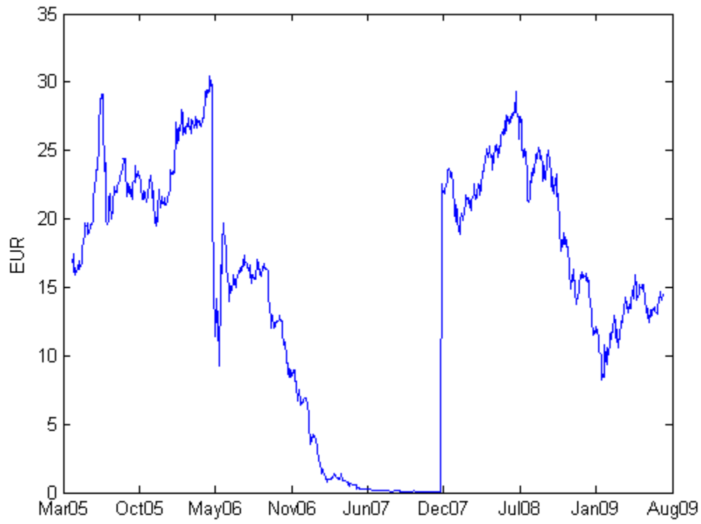

Figure 6.1: A plot of the prices of year ahead EUA futures contracts

Note the gap during mid 2006 - this was the market collapse observed during the end of phase I. As discussed previously, since EUAs from phase I weren't bankable into phase II there was still faith in the phase II market. Thus, the 2008 EUA future (the first year-end future in phase II) still maintained high prices. For our data, we will therefore let the 2008 EUA future be representative of EUA prices for the entire phase I and the first year of phase II. This concatenated data series without the gap is plotted in figure 6.2.

Table 6.2 below summarizes the descriptive statistics of EUA price returns, figure 6.3 shows a histogram of EUA returns, and figure 6.4 gives us an indication as to the distribution of EUA returns data.
From this we see that the returns data of our selected EUA futures contracts are relatively symmetric and approximately t_5 distributed.

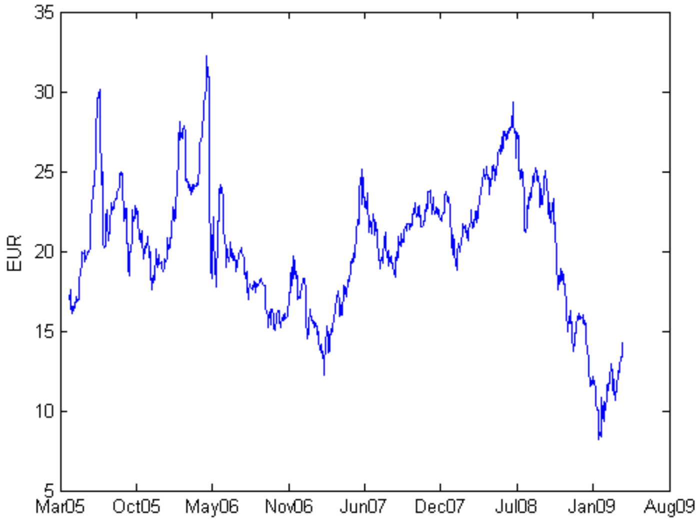

Figure 6.2: A plot of the chosen EUA data series

Std. Dev	Excess Skewness	Excess Kurtosis
0.0283	-0.1604	7.6807

Table 6.2: A table of descriptive statistics for the selected EUA futures returns data

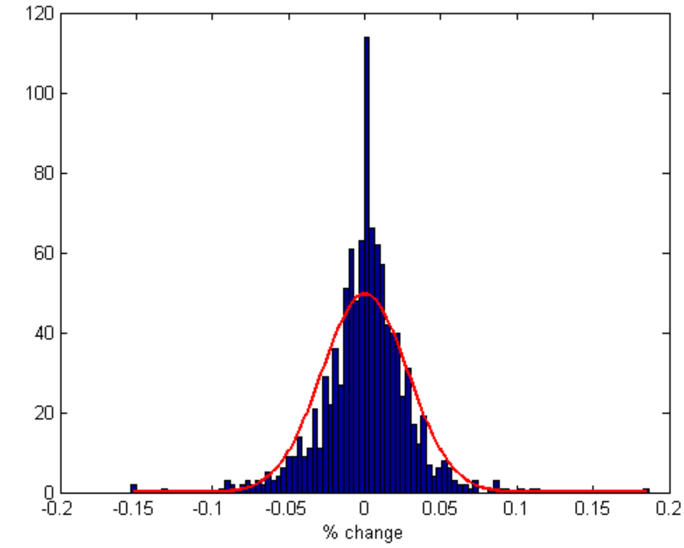

Figure 6.3: Histogram of EUA futures returns with fitted normal PDF

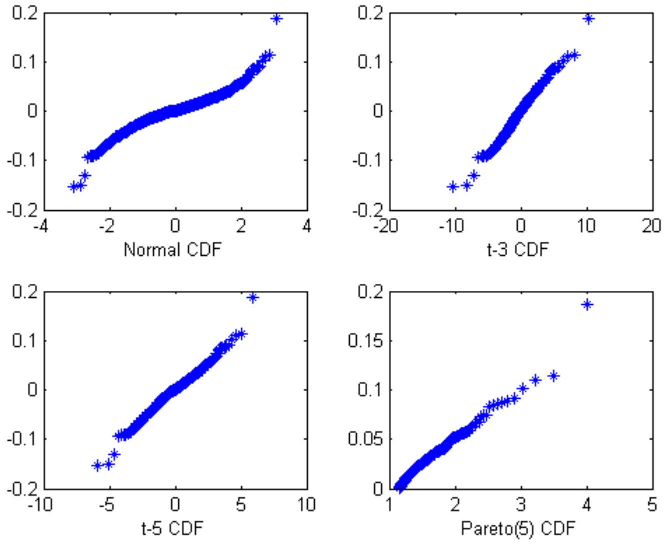

Figure 6.4: QQ plots of EUA futures returns

6.3 Other data

We will not delve deep into the other data series that we have selected, and will be content plotting the data series and examining box plots. Note that all prices have been converted to Euro, and for power, coal, and natural gas prices are stated per MWh equivalent. Coal prices (quoted as \$/ton) have been converted to €/MWh using the USDEUR spot rate and a conversion factor of 0.12286 MWh/ton coal, and natural gas prices (quoted as GBpence/therm) have been converted to €/MWh using the GBPEUR spot rate and a conversion factor of 0.02931 MWh/therm gas. Furthermore, in concatenating time series data for subsequent year ahead futures, the connection point represents a discontinuity, and the corresponding return data point is removed.

Figure 6.5 shows the original data series, and figure 6.6 shows the corresponding box-plots. We note that returns are generally not very skewed, but that there are a large number of outliers that could bias our results.

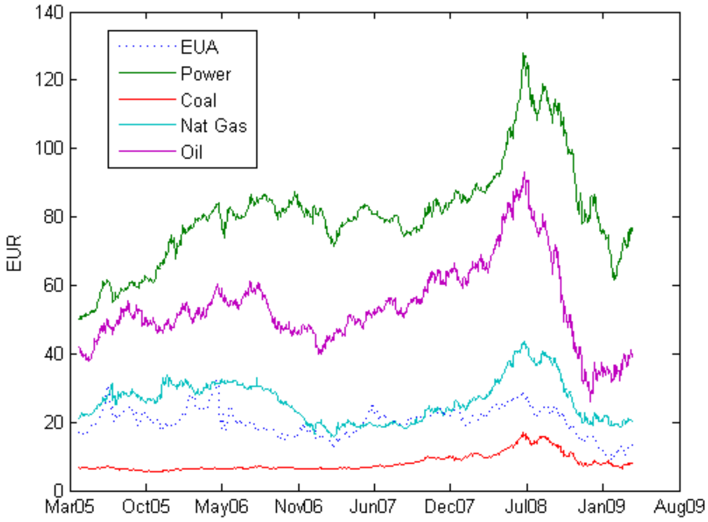

Figure 6.5: Plots of other data series

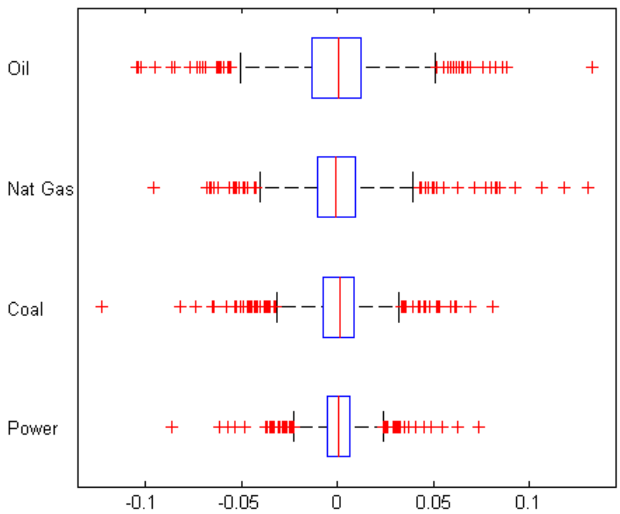

Figure 6.6: Box plots of other data series

Chapter 7

Analysis

Having discussed the mathematical tools as well as the data that will be used, we are now ready to proceed with the analysis. To begin with, the fundamentals approach to valuing carbon credits is revisited and examined for price signals. After that, this chapter is organized much like the mathematics chapter - pairwise correlations, multivariate regression, and cointegration are examined in order.

7.1 Fundamentals theory revisited

The first step is to look at dark and spark spreads. A plot of the dark and spark spreads for the period 2005-2009 is given in figure 7.1. The thermal efficiencies ρ_{coal} and ρ_{gas} are set to 36% and 50% respectively, and are widely used industry averages.

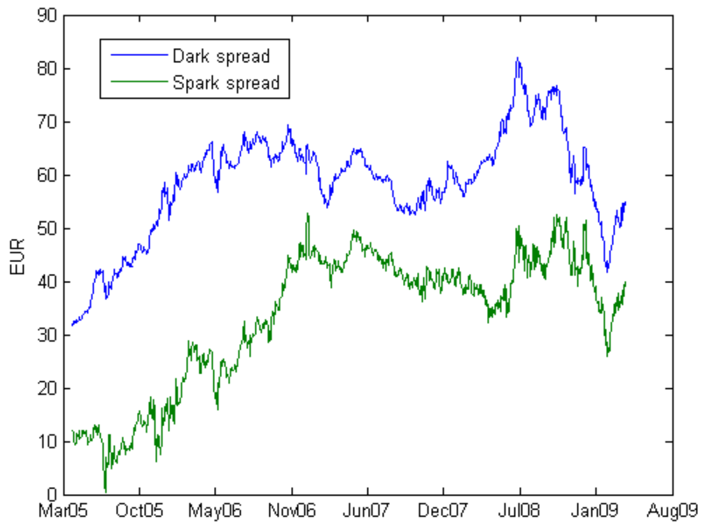

Figure 7.1: Year ahead dark and spark spreads on the German power market

Using this data, these plots show that gas fired plants are generally less lucrative than coal fired plants when polluting is free. Both spreads are positive meaning that, disregarding operational costs, both types of power production are profitable. Since gas was relatively more expensive than coal during 2005-2006, the spread was higher during this period.

Factoring in the price of carbon leads to the clean dark and spark spreads. Figure 7.2 shows the clean dark and spark spreads for the period 2005-2009. For these plots the emissions factors E_{coal} and E_{gas} are set to 0.86 tCO_2/MWh and 0.36 tCO_2/MWh, respectively.

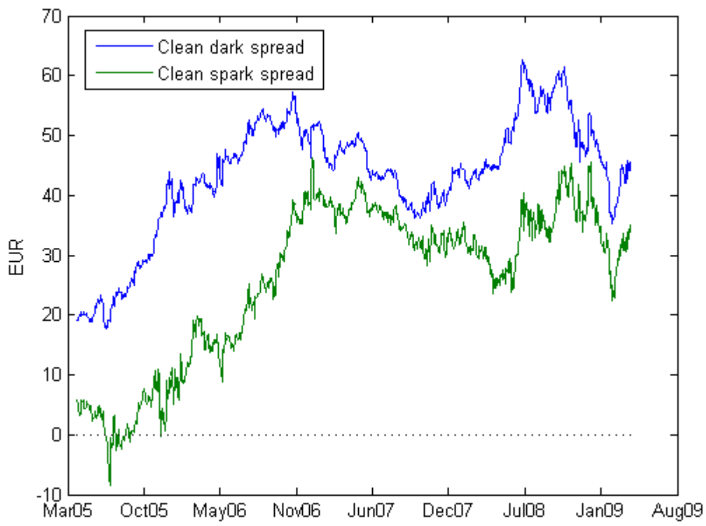

Figure 7.2: Year ahead clean dark and spark spreads on the German power market

From this plot it is seen that in a carbon constrained economy it is still generally more profitable to run a coal fired power plant than a gas fired one. The difference, however, is significantly reduced.

Finally, the implicit switching price can be calculated from the spreads above. Figure 7.3 is a plot of the switching price. Considering the clean dark and spark spreads, it is not entirely surprising that the switching price is positive and higher than EUA price. The huge differences during significant periods of time, however, are surprising. During most of 2007 it could be argued that, while it was not more economic to produce natural gas fired power, there was at least a degree of equality which may have led to higher gas fired power generation, and thereby lower total emissions.

The question at hand is whether switching price reflects actual carbon credit prices well. A glance at the plot reveals that switching price is a very poor indicator of EUA price. With different data sets and sources, this plot will look a little bit different. In any case, we see that switching price is of little help in determining a fundamental value of EUAs. EUA carbon credits must therefore be seen as an independent asset class, and are not priced according to the fundamentals theory.

Another observation is that EUA prices are relatively stable compared with the fluctuating switching price. One reason for this is that switching price is much too simple a

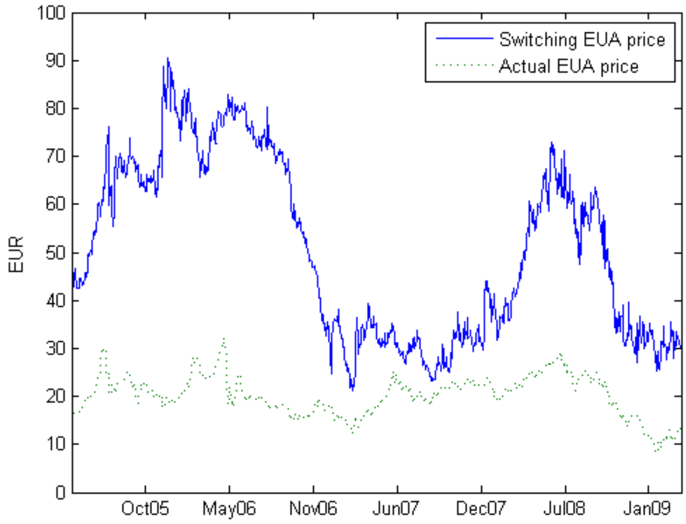

Figure 7.3: Implicit EUA switch price necessary for fuel switching from coal to gas fired power generation

model for interpreting such a complex market as carbon credits. There are many more market forces at play which can severely impact price, than what are presented in this model.

7.2 Correlation analysis

Having examined switching price as an indicator of EUA price and found it to be rather poor, we proceed with more advanced analyses in the search for market structure.

From the mathematical presentation, we know that the correlation coefficient is the determining factor of the univariate regression. We will begin by looking at pure correlations, and will extend this analysis in the next section on multivariate regressions. Table 7.1 gives the correlation matrix of the relevant data for the full dataset using standard Pearson correlation.

	EUA	Power	Nat Gas	Oil	Coal	DAX	Switch
EUA	1	0.56	0.00	0.21	0.25	0.17	-0.11
Power		1	0.03	0.25	0.48	0.19	-0.19
Nat Gas			1	0.15	0.14	0.03	0.86
Oil				1	0.33	0.26	0.00
Coal					1	0.24	-0.32
DAX						1	-0.09
Switch							1

Table 7.1: Correlation matrix for full data set

Correlations are thus not particularly high, especially considering that $R^2 = \rho^2$ is the total variability explained. It is really only power that has a significant correlation to EUA prices. The fundamentals theory states that there should be a direct link between EUA price and power price. The correlation can therefore even be considered surprisingly low. But, as we have seen, the switching price is not a particularly instructive metric - it even has a negative correlation with EUA price.

Furthermore, the table reveals that natural gas has virtually no correlation with EUAs. Also, since pairwise correlations between variables other than EUAs are quite high, we note that there is a high degree of multicollinearity in the data as a whole. We will return to this when performing multivariate regressions in the next section.

Since a full data set analysis does not reveal much with regard to the correlations to EUA returns, a moving window approach is applied to see if there are significant correlations present during short time intervals. A 60 day moving window of exponentially weighted moving average (EWMA) correlations is calculated. Shortening the sample data increases the confidence interval for the correlation coefficient. At the same time, however, having too long data series wouldn't reveal the dynamic behavior of the correlation (if such is present). 60 days is chosen as a compromise. Figures 7.4 and 7.5 show 60 day moving window correlations calculated against EUA prices.

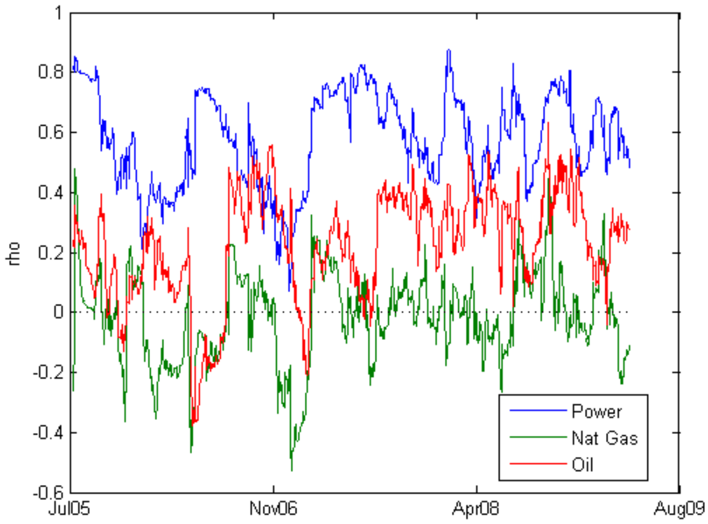

Figure 7.4: 60 day moving window correlations against EUA returns

It is seen that power and sometimes oil are the only variables with significant correlation, and during H2 2008 and H1 2009 coal was somewhat correlated. Natural gas and the DAX index show a very weak (but noisy) correlation. We conclude that correlation, both full period and EWMA, reveal little other than that power is correlated with EUAs. The lack of correlation is not entirely uninteresting, however, since it tells us that correlation is perhaps not the best tool for detecting and measuring links between these chosen energy complex variables.

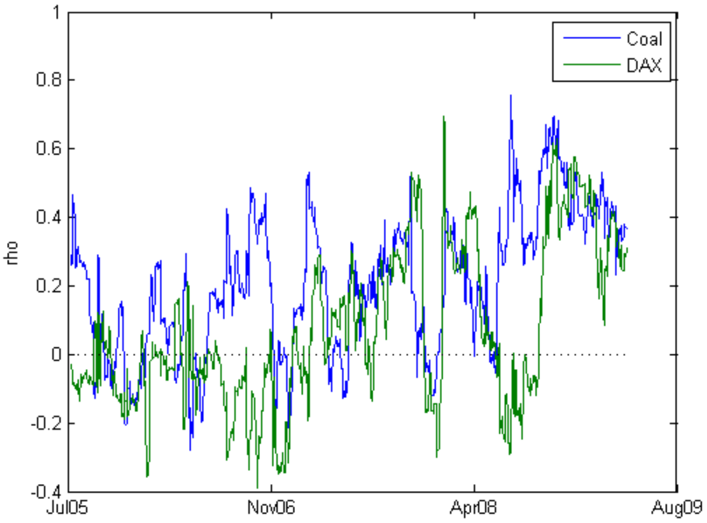

Figure 7.5: 60 day moving window correlations against EUA returns

7.3 Multivariate regression analysis

Having looked at pairwise correlations, we will now extend the analysis to multivariate regression. We will examine the degree to which a linear combination of multiple independent variables (power, nat gas, oil, coal, DAX) can explain the dependent variable (EUA returns), and draw conclusions from this. As a final analysis, the multivariate regression deemed to be the 'best' is used for a simple prediction analysis in order to see if any further conclusions can be drawn.

7.3.1 Regression

We begin by performing multivariate regression using all variables as independent variables. Due to the level of correlations that was shown to exist in the correlation analysis, we will use PCA regression in the following in order to obtain robust results despite the multicollinearity. The regressions will be performed on the full data set. The table below shows the regression coefficients (the β values) and a 95% confidence interval of the regression of all variables with respect to EUA price.

	Const	Power	Nat Gas	Oil	Coal	Dax
β	-0.0002	1.1251	-0.0245	0.0911	-0.0975	0.0954
$\beta_{0.95+}$	0.001	1.24	0.05	0.16	0.005	0.20
$\beta_{0.95-}$	-0.002	1.01	-0.10	0.03	-0.20	-0.003

For this regression $R^2 = 0.3198$ and $R^2_{adj} = 0.3165$. We see that power and oil are the only variables with significant coefficients at the 95% significance level.

We will now eliminate variables in order to maximize R^2_{adj}. The maximum is obtained when only natural gas is removed. The regression coefficient for natural gas is very low, and the correlation in the previous section was seen to be insignificant. The following table shows the regression results with natural gas removed.

	Const	Power	Oil	Coal	Dax
β	-0.0002	1.1270	0.0884	-0.1012	0.0962
$\beta_{0.95+}$	0.001	1.24	0.15	0.00	0.19
$\beta_{0.95-}$	-0.002	1.01	0.02	-0.20	-0.002

In this case $R^2 = 0.0.3195$, $R^2_{adj} = 0.3168$ for the full data set. As expected R^2 is lower, but the important metric is R^2_{adj} which is indeed somewhat higher. Since natural gas had insignificant correlation with EUAs, and was detrimental to the multivariate regression, we can safely rule out natural gas as a variable that is linked to EUAs in terms of correlation.

Since the correlation coefficient for power is high, we try regressing with only power and constant. The result is $R^2 = 0.3100$, $R^2_{adj} = 0.3094$, and the correlation coefficients are given in the table below. Although all other variables are removed, the result is only a small reduction in R^2_{adj}. We therefore conclude that power is by far the most telling variable in terms of correlation, although the correlation is still low.

	Const	Power
β	-0.0002	1.1267
$\beta_{0.95+}$	0.001	1.23
$\beta_{0.95-}$	-0.002	1.03

7.3.2 Prediction

As a further analysis using multivariate regression we will attempt to use our R^2_{adj} maximizing result to perform price prediction on in-sample data, i.e. the regression with all variables except natural gas, as shown in the previous section. While it may be considered false to perform in-sample prediction, a deep prediction analysis is beyond the scope of this paper as we are only interested in analyzing market data for structure. For in-sample data, figure 7.6 shows the predicted versus actual EUA price. The results are quite positive, and prediction is seen to work surprisingly well on in-sample data, despite correlations being low.

Here it makes sense to consider the Bayesian information criterion (BIC, see mathematics section) in order to choose the variables that are the 'best' for prediction. When using all variables, we obtain $BIC = 2.9997 \cdot 10^3$. Using only power, we again obtain the BIC minimizing result with $BIC = 2.7036 \cdot 10^3$. Even considering BIC, we can conclude that power price is the only useful variable in terms of correlation. When using all other variables except power, we get $BIC = 2.8104 \cdot 10^3$ which is in between, but the plot of predicted versus actual EUA price is rather poor as seen in figure 7.7.

As a final comment, recall that linear regression applied to financial data is prone to regime switching. A multivariate linear relationship that has been found to hold may be short-lived. A full prediction analysis should calibrate the model to recent historical data and measure the predictive power on future out-of-sample data.

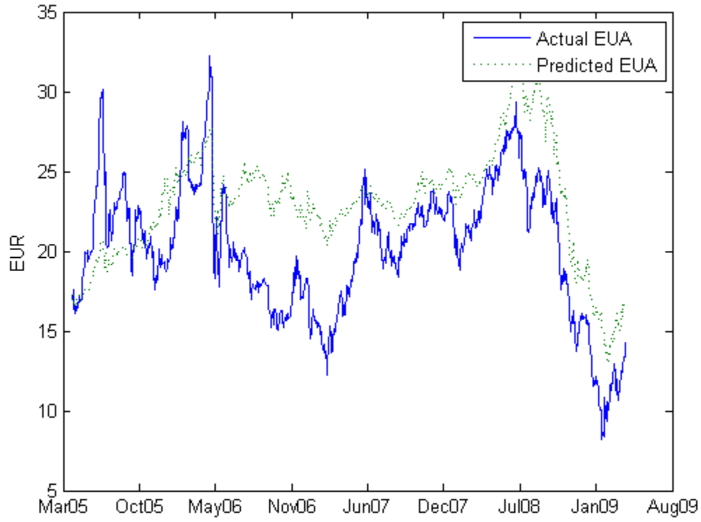

Figure 7.6: Multivariate regression prediction using all variables

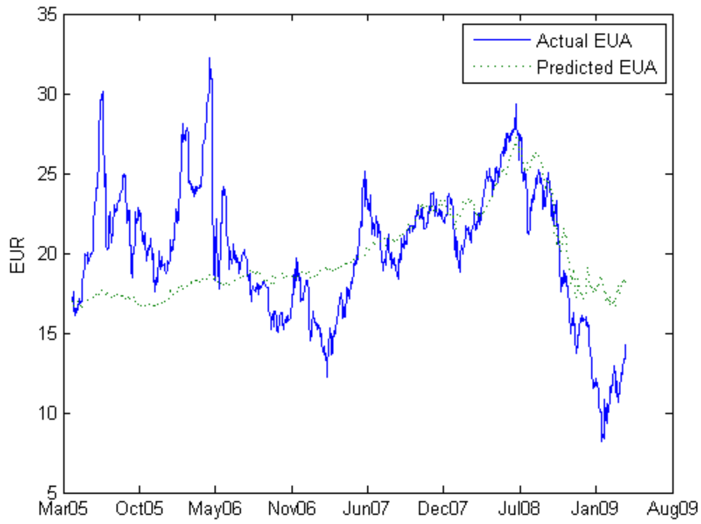

Figure 7.7: Multivariate regression prediction using all variables except power

7.3.3 Regression conclusions

What we have been able to conclude using multivariate regression in this section is that power is the only variable with a significant link to EUA prices in terms of correlation. Natural gas is found to be insignificant in terms of correlation. This is not surprising given market participants' strong belief in the fundamentals theory.[1] One possible explanation for power price's dominance as a signal is that natural gas and coal markets are much less liquid than the power market.

Even with correction for multicorrelation using PCA, the results of the regression analysis were quite poor. There may be several reasons for this. First and foremost, it is appropriate to question whether the data should obey a linear relationship at all, i.e. whether data is 'generated' according to a linear relationship. Furthermore, even if a linear relationship is appropriate, the assets may not be linked on the daily-closing basis that would be necessary to cause correlation and thus meaningful regression. Since correlation based analysis did not reveal much, the natural question is whether there is any longer term relationship between the assets. This is the topic of the next section.

A final comment is that no lagged variables have been included in the regression above. The reason is that, in a separate analysis, the coefficients of the lagged variables were found to be insignificant. This was also reflected in the fact that residuals for the standard multivariate regression with all variables was found to be white noise. This means that little or no further information can be extracted using variable lags.

7.4 Cointegration - truths revealed

A quick recap of cointegration as explained in the mathematics section tells us that we must test for stationarity of the residual (the difference) between two logged and scaled time series. The hypothesis in the Augmented Dickey-Fuller (ADF) test of stationarity is $H_0 : \beta = 0, H_1 : \beta < 0$. Not rejecting the null hypothesis (at a certain significance level) is equivalent to the time series having a unit root. This means that the difference between the scaled time series is non-stationary, and no cointegrating relationship exists.

Table 7.2 shows a cointegration hypothesis-test matrix. A '1' indicates that the null hypothesis is rejected at the 5% significance level, meaning that there is a statistically significant cointegrated relationship between the two variables. The number in parenthesis indicates the t-score of the ADF test for those variables where the null hypothesis is rejected. At the 5% significance level the critical value is given by $t_{critical} = -2.865$, and at the 1% significance level $t_{critical} = -3.4378$. Finally, since cointegration analysis is able to handle infrequently sampled data sets, total industrial output has been included as a variable as it may reveal interesting results.

To begin with, we examine relationships with regard to EUAs. Natural gas and oil are both cointegrated with EUAs at the 5% level, but only oil is cointegrated at the 1% level. The regression residuals for the pairwise relationships are plotted in figure 7.8.

These plots do seem to exhibit stationarity and mean reversion. This shows that even though natural gas and, to a large extent, oil had low correlation with EUAs, they do

[1] Sikorski (2009)

	EUA	Power	Nat Gas	Oil	Coal	DAX	Ind. Prod.
EUA		0	1 (-2.94)	1 (-3.77)	0	0	0
Power			0	0	1 (-3.24)	0	0
Nat Gas				0	0	0	0
Oil					0	0	0
Coal						0	0
Dax							1 (-2.98)
Ind. Prod.							

Table 7.2: Cointegration hypothesis-test matrix

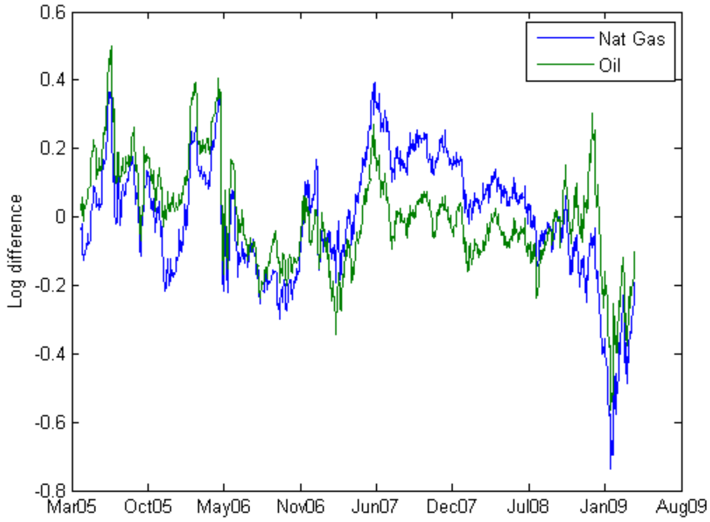

Figure 7.8: Pairwise residuals for natural gas and oil to EUA

exhibit a cointegrating relationship. This is an example of the situation of uncorrelated but co-trending time series, as illustrated in the mathematics section. The conclusion of this is that, in attempting to predict future EUA prices, oil and natural gas are the determining variables. Though its own asset class, EUAs are indeed shown to be linked to these commodities. Oil is the leading proxy for energy consumption, and it is therefore not entirely surprising that EUAs should be linked to this proxy in the long run.

In continuing the analysis, it is very interesting to note that, although EUAs and power are highly correlated, there is no cointegrated relationship. Although daily closing prices move more or less in sync, there is nothing to suggest that prices follow each other. This is an example of correlated but divergent time series. A plot of the residuals is shown in figure 7.9, and does indeed look non-stationary.

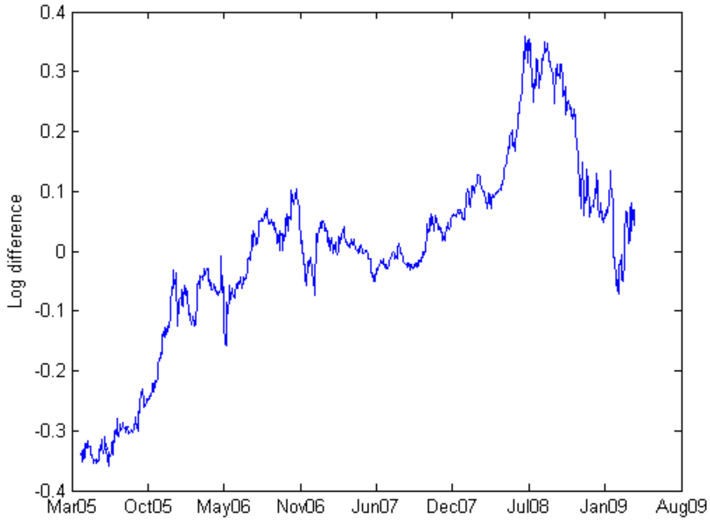

Figure 7.9: Residuals for power and EUA

The fact that power and coal are cointegrated is noteworthy since it means that, in terms of long run equilibrium, power and coal are linked, and both are disconnected from the separate carbon credit 'complex' consisting of EUAs, natural gas, and oil. It is easy to speculate that there is a causality between coal and power, i.e. that coal prices to a large extent dictate power prices.

Finally, it is less surprising that the DAX and industrial production indexes are cointegrated since the DAX to a large degree represents industrial production.

A critique of the analyses above is that H2 2008 and H1 2009 data represent points during the recent financial crisis. It is well known that during crises, many assets become highly correlated. From the moving window correlation plot in the correlation section, there was no reason to believe that correlations had increased significantly. This, however, does not preclude that assets become highly cointegrated, which could bias the above results. At the worst, it has been shown here that there is insight to be gained from performing a cointegration analysis.

Chapter 8

Summary and conclusions

This report is a step in the direction of understanding the market forces active in the EU ETS, and, more specifically, the carbon constrained energy complex. Through empirical statistical analyses, we have been able to draw several conclusions as to the nature of these forces, as well as to the applicability of various statistical methods to energy complex data in general. We have, in order, examined the fundamentals theory, correlation, multivariate regression, and cointegration in the German energy complex.

The fundamentals theory dictates that switching price should be of relevance in determining the price of EUAs. It was found, however, that switching price is a very poor indicator of EUA price, and that the fundamentals theory is a far too simple mechanism to capture the complex workings of this new asset class.

Next, since many authors before believed correlation to be an appropriate tool in the context of the energy complex, we examined both pure correlation and multivariate regression. Contrary to several authors, however, we find that hardly any significant correlations and linear relationships exist. The analysis revealed that correlations with EUAs are in general very low for most chosen assets. Power was the only asset shown to have any substantial correlation. Multivariate regression mostly reinforced the conclusions of the correlation analysis.

Cointegration applied to EUAs is, however, relatively unexplored territory. Being a tool for examining longer term links, and not just the degree to which daily closing prices move in sync, as is the case with correlation, it revealed much more of the market structure. EUAs were found to be cointegrated with both natural gas and oil, and, surprisingly, not at all with power.

We have thus picked apart what variables are important, and why. In the short run, day-to-day closing prices of EUAs are linked to power prices, but in the long run EUAs are cointegrated with natural gas and oil.

A suggestion for future work is to conduct a similar analysis, but with UK market data using UK relevant commodities, or, alternatively, conducting an EU wide analysis using EU wide indexes. Further it would be interesting to examine optimal hedging of EUAs (or other energy complex commodities) with a basket of commodities using cointegration rather than correlation, as is common practice.

The EU ETS is a dynamic system that may have an entirely different structure years from now. The up-coming climate summit in Copenhagen in December 2009 may drastically alter the system's future outlook. It will therefore be interesting to see how the market forces change with the evolution and proliferation of cap-and-trade systems around the globe - in the US, Japan, Canada, Australia and other countries - and how these will influence each other and possibly become linked.

As with all free markets, however, asset prices appear chaotic, and it will always be necessary to pick apart market data in order to build an understanding for their functioning. The exploration of statistical tools and approaches in this report is thus a contribution to market practitioners seeking a deeper understanding.

Appendix

Generating correlated discrete-time geometric Brownian motions

Consider two discrete-time fictional assets S^1 and S^2 that are defined as discrete geometric Brownian motions,

$$dS_t^1/S_t^1 = \mu_1 dt + \sigma_1 dB_t^1, S_0^1 = S^1(0)$$

$$dS_t^1/S_t^1 = \mu_2 dt + \sigma_2 dB_t^2, S_0^2 = S^2(0)$$

We know that it is the correlation between the Brownian motions dB^1 and dB^2 that defines the correlation of the price processes. When simulating discrete price processes, it is possible to hard-code the correlation between the two by simulating $\sigma_1^2 \delta B^1$ and $\sigma_2^2 \delta B^2$ that have a fixed correlation $\rho_{1,2}$. Let Σ denote the covariance matrix,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

We can then use the fact that multiplying a 2×1 vector of two independent $N(0,1)$ variables by the lower triangular Cholesky factorization of Σ, gives us correlated normally distributed Brownian motion increments multiplied by their respective standard deviations. Mathematically this is given by

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim N\left(0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

$$Chol(\Sigma) = A^T A$$

$$\begin{bmatrix} \sigma_1 \delta B_1 \\ \sigma_2 \delta B_2 \end{bmatrix} = A \cdot \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

The reason that this works is as follows. If $\mathbf{Z} \sim N(0, \mathbf{I})$ then $C^T\mathbf{Z} \sim N(0, C^T C)$. Thus, our problem reduces to finding C such that $\Sigma = C^T C$. Since Σ is symmetric and positive definite, it can be written as $\Sigma = U^T D U$ where U is an upper diagonal matrix, and D is a diagonal matrix. From this it follows that

$$\begin{aligned} \Sigma &= U^T D U \\ &= (U^T \sqrt{D})(\sqrt{D} U) \\ &= (\sqrt{D} U^T)(\sqrt{D} U) \\ &= A^T A \end{aligned}$$

Principal component analysis

Principal component analysis has many uses, and is based on the idea of transforming correlated data series into uncorrelated data series - i.e. they are orthogonalized. Each 'principal component' refers one transformed data series. Assuming $\mathbf{X}$ is a matrix whose column vectors are the normed original data, the correlation matrix is given by $\mathbf{V} = \mathbf{X^T X/T}$. Let $\mathbf{W}$ represent the matrix containing the eigenvectors of $\mathbf{V}$, and $\mathbf{\Lambda}$ represent the matrix with the eigenvalues in the diagonal. By the definition of eigenvectors and eigenvalues it is easy to see that

$$\mathbf{VW} = \mathbf{W\Lambda}$$

Now let the m columns in $\mathbf{W}$ be ordered such that column w_i corresponds to eigenvalue λ_i, and $\lambda_1 > \lambda_2 > \ldots > \lambda_m$. Call this new ordered matrix $\widetilde{W}$. The ith principal component of the system is given by

$$P_i = \mathbf{Xw_i}$$

This results in a matrix $\mathbf{P}$ with the principal components as column vectors, and is given by

$$\mathbf{P} = \mathbf{X\widetilde{W}}$$

$\mathbf{P}$ is the matrix of transformed uncorrelated data series as its column vectors. It is easy to prove that the matrix $\mathbf{P}$ is orthogonal

$$\begin{aligned}
\mathbf{P^T P} &= \mathbf{\widetilde{W}^T X X \widetilde{W}} \\
&= T\mathbf{\widetilde{W}^T \widetilde{W} \Lambda} \\
&= (\mathbf{\widetilde{W}\ is\ orthogonal} \Rightarrow \mathbf{\widetilde{W}^T} = \mathbf{\widetilde{W}^{-1}}) \\
&= T\mathbf{\Lambda}
\end{aligned}$$

which is diagonal. The proportion of the total variation in $\mathbf{X}$ that is explained by P_m is

$$\frac{\lambda_m}{\sum \lambda_i} = \frac{\lambda_m}{k} \ (tr(\mathbf{\Lambda}) = \mathbf{tr}(\mathbf{V}) = \mathbf{k})$$

In a highly correlated system $\lambda_1 >> \lambda_2 > \lambda_3 > \ldots$, meaning that the first principal component alone will explain much of the variation in the system. Finally, an essential aspect of PCA is that the original data matrix $\mathbf{X}$ can be written as a linear combination of the principal components as

$$\mathbf{X} = \mathbf{P\widetilde{W}^T}$$

This is the so called 'principal component representation' of the original data series.

Bibliography

[1] Aggeryd, J. & Strömqvist F. (2008). *An Empirical Examination of the EUA Emission Rights Market.* Stockholm School of Economics.

[2] Alexander, C. (2001). *Market Models: A Guide to Financial Data Analysis.* John Wiley & Sons Ltd, West Sussex.

[3] Benz, E. & Truck, S. (2007). *Modeling the Price Dynamics of CO_2 Emission Allowances.* Working Paper Bonn Graduate School of Economics, Germany.

[4] Daskalaki, G. & Psychoyios, D. & Markellos, R (2008). *Modelling CO_2 Emission Allowance Prices and Derivatives: Evidence from the European Trading Scheme.* Journal of Banking & Finance, 2009.

[5] Gardner, S. (March 2008). *EU ETS – Saddle up for lower emissions.* Ethical Corporation. `http://www.ethicalcorp.com/content.asp?ContentID=5786`.

[6] Kanamura, T. (2009). *A Classification Study of Carbon Assets into Commodities.* Working Paper, J-Power.

[7] *Möjligheter och kostnader för att reducera växthusgasutsläpp i Sverige.* (2008). McKinsey&Company, Stockholm.

[8] POMAR (2007).*Market Analysis and Risk Management of EU Emissions Trading.* University of Helsinki & Helsinki University of Technology.

[9] Paolella, M. & Taschini, L. (2006). *An Econometric Analysis of Emission Trading Allowances.* Swiss Finance Institute Research Paper No. 06-26.

[10] Mansanet-Bataller, M. & Tornero, A. & Micó, E. (2006). *CO_2 Prices, Energy and Weather.* Working Paper, Department of Financial Economics, University of Valencia.

[11] Mansanet-Bataller, M. & Pardo, A. (2008). *CO_2 Prices and Portfolio Management.* Working Paper.

[12] Sikorski, T (2009). *Monthly Carbon Standard, 6 March 2009.* Barclays Capital, London.

[13] *United Nations Framework Convention on Climate Change.* `http://unfccc.int/essential_background/items/2877.php`