Master Thesis in Quantitative Finance

Investing in Commodity Indices: Risk Factors and Trading Strategies

Academic supervisor
Karolina Scholtus
Erasmus University of Rotterdam

External supervisor
Tjeerd van Cappelle
ING Investment Management

Co-reader
Dr. Erik Kole
Erasmus University of Rotterdam

Sirvydas Dagys
Erasmus University of Rotterdam
Erasmus School of Economics
Student number 346351

September 18, 2011

Abstract

This paper researches risk factors and trading strategies using 27 Dow Jones-UBS commodity indices. A stepwise regression procedure with several criteria is established that leads to the final time-series risk model choice. It consists of sector, liquidity, materials sector equity, momentum and non-commercial hedging pressure factors. The model has an average out-of-sample R^2 of 0.57.

In the strategy part of the research, two types of framework for every strategy are considered. In one the strategy goes long/short in the sectors, in the other it goes long/short in individual commodities within each sector. The Henriksson-Merton non-parametric test is used to evaluate strategies. The strategy based on the roll yield deviation from its' 5-year average is the best across sectors with the IR of 0.54 and the 12-month momentum with open interest growth as confirmation is chosen within sectors with the IR of 0.28. The strategies have no structural breaks in the sample period and are robust to the choice of parameters. As the two strategies have low correlation because of different construction framework, the combination of them can reduce risk.

Keywords: Commodity indices; Commodity risk factors; Trading strategies; Momentum; Roll yield; Henriksson-Merton test.

Contents

List of Tables

List of Figures

Preface

This study is the Master thesis work for the Quantitative Finance programme at Erasmus University of Rotterdam. It has been carried out as a research internship in the Strategic and Tactical Asset Allocation Group (STAAG) at ING Investment Management (ING IM). ING IM is a global asset manager within the ING Group with €326 billion[1] in assets under management. STAAG is a boutique that manages a variety of multi-asset and multi-sector portfolios. This setting makes the thesis to be both practically relevant and academically thorough and challenging. I thank my supervisor at ING IM Tjeerd van Cappelle for all the ideas and support during the internship and my academic supervisor Karolina Scholtus who was very active in discussing and providing feedback. I am also grateful for Valentijn van Nieuwenhuijzen and Koen Straetmans for their comments throughout the research.

[1]Data as of June, 2011.

1 Introduction

The research is conducted at ING IM, therefore it is practically oriented and all results can be directly implemented in practice. In addition, the econometric techniques and decision criteria we use can be of interest to academics as well as practitioners.

The paper is aimed at two major issues, namely establishing a risk model and finding successful trading strategies for commodities. The risk factor model is generally used to reduce dimensions in various problems. It is easier to deal with several factors than with many assets in a portfolio. For example, when constructing an optimal portfolio, it is easier to estimate the factor covariance matrix instead of the covariance matrix of all assets. In risk management it plays a big role as well as one can track and hedge the exposure to certain factors. This way investor managers can isolate themselves from risks they or their clients are not comfortable with. It is important that the factors are backed theoretically as well as empirically. The strategy research builds upon risk model analysis in the sense that it provides the basis to develop strategies once we have a sense what drives the commodity returns from the first part of the study. Economic rationale and empirical evidence is demanded trading strategies as well. First, there has to be a theory why a strategy should work and then it is tested to see if risk-adjusted excess returns can be achieved. In general, the strategies try to identify situations where various agents participating in the market are willing to pay a premium.

Investments in commodities and the number of related papers have been growing in recent years (Tang and Xiong, 2010; Fuertes et al., 2010b; Vrugt et al., 2007), however, commodities are still under-researched compared to other asset classes. The existing literature usually looks into a few theories at the time but there is a lack of research that carries out an analysis on a wide set of factors or strategies. Therefore, it is hard to say what the interaction among different factors is. The thesis analyzes an extensive list of factors and strategies including both compiled from the existing literature and proposed new ones. Also, most of the existing literature focuses on investing in commodity futures contracts but some of the investors might have restricted access to futures contracts. We add to the literature by considering strategies based on commodity indices.

The research is based on 27 Dow Jones-UBS commodity total return indices that span across five sectors: energy, precious metals, industrial metals, agriculture and livestock. The first part of the thesis deals with the factors that explain the variance in commodity returns. We consider both time-series and cross-sectional models for the univariate analysis and continue to the multivariate analysis with the better one. There is a stepwise regression procedure based on several criteria established that chooses the final model. The final time-series model consists of sector, liquidity, materials sector equity, 3-month momentum and non-commercial hedging pressure factors and has an average out-of-sample R^2 of 0.57. It does not pass the Gibbons-Ross-Shanken (GRS) (see Gibbons et al., 1989) asset pricing test at 5% significance level but we list several issues that are of consideration and propose ideas for further research that could tackle them. The risk model research is a good basis for the second part of this thesis that deals with trading strategies. The theory behind some of the risk factors and knowledge gained can be effectively used in constructing strategies.

There is a list of strategies tested and all of them take balanced long/short positions in two types of set up: in the first case long/short positions are taken on sector level and in the second case long/short positions are taken within every sector thus remaining sector neutral. In each of these types we select the best strategy according to the Henriksson-Merton (HM) non-parametric test (see Henriksson and Merton, 1981). The best strategy that invests in sectors is based on a roll yield deviation from its' 5-year average and has a significant HM test statistic p-value of 0.004 and Information Ratio of 0.543. 12-month momentum with open interest growth as confirmation performs the best according to our criteria within the sectors. It has the HM p-value of 0.005 and IR of 0.277. As these strategies are weakly correlated, we show that a combination of them reduces risk. Finally, we carry out structural change tests and robustness checks.

The paper is organized as follows. Section 2 provides an overview of the current literature on investing in commodities. The data, constructed variables and sources used in this research are described in Section 3. Section 4 outlines some general summary statistics. The next two sections provide the methodology and results for the risk model estimation. In Section 7 the GRS asset pricing test is performed on the estimated model. Section 8 introduces the methodology for the construction and testing of trading strategies together with the results. The next section tests for possible structural changes over time in the chosen strategies. Section 10 performs some robustness checks on the arbitrary parameters in the strategies. Section 11 concludes and provides ideas for further research.

2 Literature overview

In this section we review the literature that researches the drivers behind commodity returns. This overview helps us later produce lists of both factors and strategies that are tested. Most of the literature focuses on one or two commodity specific factors. Some of it also includes various macroeconomic factors or the factors known to explain stocks or bonds. However, we did not find any research that would try to find the set of factors explaining the commodity returns by aggregating all proposed variables. Therefore, we carry out a research on an extensive list of factors many of which were proposed in the literature earlier. The literature overview is thus organized by factor themes.

Two widely researched themes in commodity investing are momentum and carry. The first one is well documented by Pirrong (2005); Shen et al. (2007) who find momentum and reversals in commodity futures. Miffre and Rallis (2007) on the other hand report that only momentum strategies work. All of these papers include common factors from stock and bond returns and show that momentum in commodity futures is not explained by them. Durr and Voegeli (2009) analyze term structure based commodity investing. Fuertes et al. (2010a,b) combine momentum and carry strategies and report an improvement in terms of risk-adjusted returns.

Carry strategy is usually based on the theory of normal backwardation that dates back to Keynes (1930) and Hicks (1939). It states that risk-averse speculators earn a risk premium by being long in the futures and taking the price risk that is transferred by the hedgers. Cootner (1960) expanded on this by showing that the sign of the risk premium can vary depending if the hedging demand is mainly driven by commodity producers or consumers. When consumers are hedging by being long, speculators earn the risk premium by being short. More recent studies were carried out by Bessembinder (1992) and De Roon et al. (2000) who find support for the theory of backwardation.

Chang (1985) splits the hedging pressure hypothesis in two sub-hypotheses. One claims that speculators earn a risk premium by simply taking positions against the hedgers and the other one states that they earn a risk premium because of superior forecasting ability. He finds support for both of these hypotheses. Recently Basu and Miffre (2011) continued along similar lines and obtained risk premium by constructing factor mimicking portfolios based on hedgers' and speculators' open interest. Also, their research partly explains the returns from portfolios based on momentum and term structure. On the other hand, hedging pressure based portfolios were also associated with higher lagged volatility and decrease in inflation hedging capability. A recent article by Hong and Yogo (2011) claims that the growth rate in total open interest is a better predictor for commodity prices than the net position of hedgers. They claim that open interest growth is a better proxy for future economic activity than past prices.

A slightly different approach has been employed by Gorton et al. (2007) who claim that the risk premia offered by carry and momentum strategies stems from investing in commodities with low inventories and is explained by the theory of storage. It dates back to Kaldor (1939) and Working (1949). Gorton et al. (2007) analyzed it on an extensive set of inventories data. They show that the volatility is higher when inventory levels are low, leading to higher risk premium in commodity futures. The literature offers an explanation that high inventories can absorb the shocks to demand and supply better, therefore resulting in lower volatility in commodity prices. Gorton et al. (2007) also state that net positions of hedgers are contemporaneously correlated with inventories and futures prices but do not find evidence of these positions being correlated with ex-ante risk premiums.

Chen et al. (2010) show that several commodity currencies can predict commodity prices and attribute this to exchange rates being more forward-looking than commodity prices. Akram (2009) researches the relationship between commodity prices, US dollar and real interest rates and find that when real interest rates fall or dollar depreciates, commodity prices rise.

There is an abundance of articles involving macroeconomic factors. Bessembinder H. and Chan (1992) study Treasury yields, equity dividend yields and junk bond premium influence in futures markets as these are known to have predictive power in equity and bond markets. Vrugt et al. (2007) use a set variables that indicate business cycle, monetary environment and market sentiment to successfully forecast the S&P Goldman Sachs Commodity Index (GSCI). Several studies analyze both commodity specific and market wide explanatory variables, including Roache (2008); Hong and Yogo (2009) and Szymanowska et al. (2010). Roache (2008) uses global equity market returns, inflation shocks, real short-term interest rates, change in the slope of the yield curve and the US dollar effective exchange rate. Hong and Yogo (2009) analyze futures basis (carry), short rate, yield spread and dividend yield. Szymanowska et al. (2010) research includes futures basis, hedgers' positions (open interest), momentum, term spread, credit spread and dividend yield.

It has also been documented that some commodities exhibit seasonal patterns that have to be considered. Sørensen (2002) analyzes seasonality in agricultural commodities. He models the futures prices with a seasonal component and estimates the parameters using the Kalman filter. Sørensen (2002) finds seasonality to be an important feature for agricultural commodity futures. Richter and Sørensen (2003) show that both price and volatility have seasonal patterns using soybean futures. Borovkova and Geman (2006) estimate the seasonal forward premium for electricity, gas, gasoil and oil futures. This paper takes into account the effect of seasonality on commodity indices.

3 Data and variable definitions

In this section variables used in the research are explained. For this research we need commodity indices returns, various data for factors and a risk-free rate. The requirements for the data are that it is not forward-looking, sufficient historical series are available and the continuity of availability in the future is to a certain extent reliable. All the data is obtained from Thomson Reuters Datastream with the exception of futures prices that are taken from Bloomberg. Data series go back to January, 1991 unless stated otherwise. The data period ends in March, 2011. Both monthly and weekly frequency is considered in this research. While weekly frequency obviously provides more observations and possibly better estimates, some of the data is not available on a higher than monthly frequency. Also, the publication lags that have to be taken into account introduce larger asynchronicity of the data in weekly frequency.

ING Investment Management cannot trade in commodity futures for all mandates because of regulatory constraints, therefore, analysis is done on total return Dow Jones-UBS commodity indices that track the performance of designated contracts in commodities. For all data series in this research the returns are converted to euro returns. Where applicable, the currency exchange costs are considered to be negligible. The 27 commodities used are grouped into 5 sectors:

- Energy: Brent Crude, Gas Oil, Crude Light, Heating Oil, Natural Gas, Unleaded Gasoline

- Industrial Metals: Aluminium, Copper, Lead, Nickel, Tin, Zinc
- Precious Metals: Gold, Platinum, Silver
- Agriculture: Orange Juice, Soybean Meal, Cocoa, Coffee, Corn, Cotton, Soybean Oil, Sugar, Wheat
- Livestock: Feeder Cattle, Lean Hogs, Live Cattle

Detailed description of the construction of the indices can be found in Dow Jones Indexes (2010). The list with exact names of the indices is available in Appendix A.

There is a mistake in Thomson Reuters Datastream database for Gas Oil total return series. Up to and including 15th of October, 2010 it actually cointains excess return series. From that point onwards it is the correct total return series. Because of this we downloaded the Gas Oil total return series manually from the Dow Jones-UBS website. Small subsamples of the Datastream and DJ-UBS gas oil total return series are presented for comparison in Appendix B.

A choice has to be made for the risk-free rate. Since in this research we want the returns to be replicable practically, actual funding rates are used instead of government rates. It is assumed that they are risk-free in the sense that they are the lowest a reputable asset manager can pay for short-term funding in practice. They are also widely assumed to be risk-free by counterparties in pricing forwards and futures. For the euro denominated instruments the German Interbank 1-month rate is used while for the US dollar we use the US Euro Currency Deposits 1-month rate. All non-zero investment returns are in excess of the corresponding risk-free rate.

The equity and bond indices can be invested in either by directly buying the constituents or investing in exchange-traded funds (ETFs) that track the indices close enough. For the non-investable factors mimicking portfolios have to be constructed. Either a basket return of commodities weighted according to the z-score of the associated factor is used or a long-short portfolio of top $\frac{1}{3}$ and bottom $\frac{1}{3}$ equally weighted commodities in a similar fashion as Fama and French (1993) did for stocks. When the distribution of a variable exhibits excessive skewness, z-scores are avoided. The portfolios are constructed to be sector neutral by first producing mimicking portfolios for each sector separately and then taking the equally weighted average of the portfolio returns. The factor returns used for the risk model are contemporaneous. However, the factor mimicking portfolios are constructed using previous month fundamental values in order to be replicable in advance. For example, this month's roll yield buckets are constructed using previous month's roll yield and then this month's bucket portfolio return is calculated. Some of the variables that end up not included in the final risk model will be tested for timing capabilities. Variable construction and the description of how it is used is listed below grouped by themes. If the variable was used in the literature earlier, references are provided.

Roll yield

We use a sector neutral portfolio based on roll yield buckets. The roll yield is defined as the monthly return on rolling a 1-month contract to a 3-month contract adjusted for the interest rate effects. The expected futures prices are $F_1 = S_0 e^r$ and $F_3 = S_0 e^{3r}$ where S_0 is the spot price and r is the monthly risk-free rate. Therefore, the expected monthly roll yield is as follows:

$$\begin{aligned} \text{Expected roll yield} &= \frac{\ln(S_0 e^r) - \ln(S_0 e^{3r})}{2} \\ &= -r \end{aligned}$$

The adjusted monthly roll yield then is

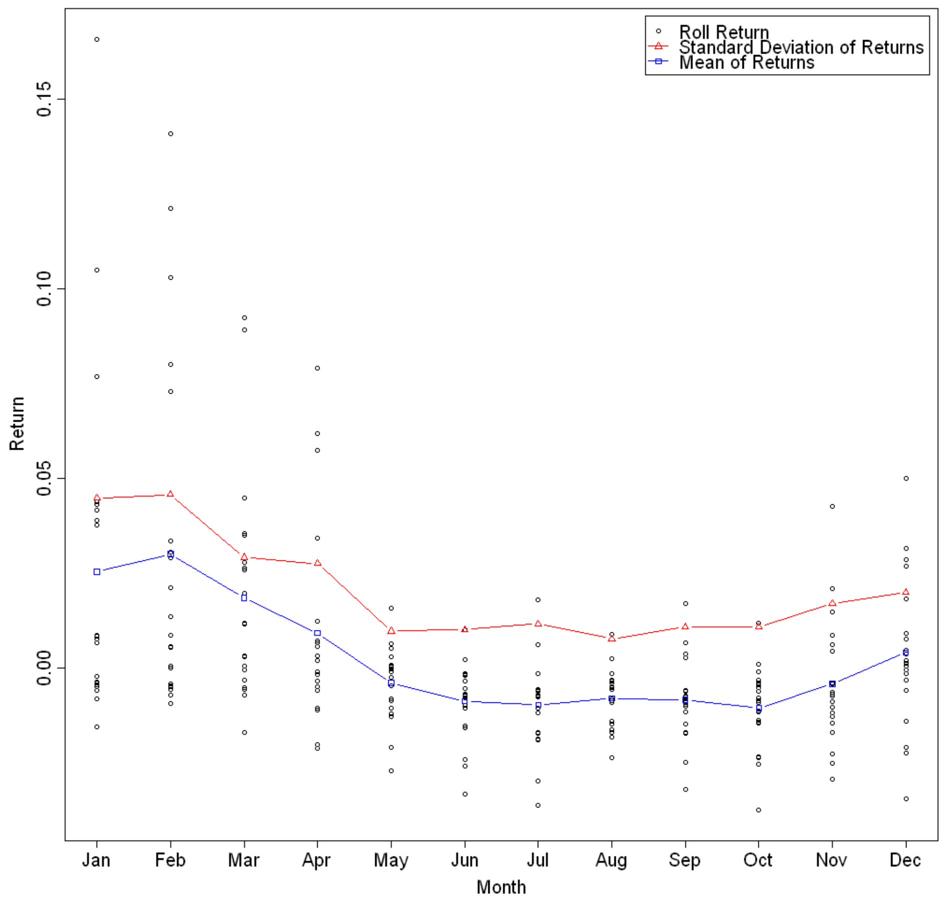
$$\text{Roll yield} = \frac{f_1 - f_3}{2} + r \tag{1}$$

where f_1 and f_3 are the logarithms of futures prices and r is the US dollar risk-free rate. The futures contracts prices are obtained from Bloomberg and we have them only for 21 commodities out of 27. Also, for some of them the data series start only in 1997.

It is known that some commodity futures prices have seasonal patterns (see Borovkova and Geman, 2006; Richter and Sørensen, 2003; Sørensen, 2002). Therefore, one can expect seasonality in roll yields as well. Indeed, for example, by looking at the roll yield graph for heating oil for each month (Figure 1), a seasonal pattern is visible.

Heating oil roll yields tend to increase during the winter months (in the northern hemisphere which represents most of the economic activity of the world) and then decrease during the spring to reach a trough in the summer. Similar pattern is observed for the volatility. To check which commodities have seasonality in roll yields, Kruskal-Wallis analysis of

Figure 1: Heating oil roll yield



variance by ranks (see Kruskal and Wallis, 1952) was performed on the roll yields grouped by months and divided by corresponding month's standard deviation (to account for differences in standard deviation). The standard deviation was estimated using an expanding window with the starting window of 10 years. The test results are presented in Table 1.

From the test results we find out that the null hypothesis of equal means is rejected for 9 roll yield series: heating oil, natural gas, corn,

Table 1: Kruskal-Wallis test results for roll yields

	Kruskal-Wallis rank sum statistic	p-value
Brent Crude	1.5572	0.9995
Gas Oil	13.2865	0.2750
Crude Light	2.9924	0.9908
Heating Oil	**60.7521**	0.0000
Natural Gas	**79.9607**	0.0000
Aluminium	8.7972	0.6406
Copper	2.1476	0.9979
Lead	8.8158	0.6389
Nickel	5.0518	0.9286
Zinc	8.5298	0.6652
Gold	4.0390	0.9688
Silver	10.1950	0.5129
Cocoa	7.4196	0.7642
Coffee	13.6365	0.2538
Corn	**70.2990**	0.0000
Cotton	**26.5016**	0.0055
Sugar	**20.7983**	0.0355
Wheat	**71.3057**	0.0000
Feeder Cattle	**41.2960**	0.0000
Lean Hogs	**71.7972**	0.0000
Live Cattle	**53.5952**	0.0000

Kruskal-Wallis analysis of variance by ranks results for commodity roll yields. The significant (with a 5% significance level) test statistics are in bold.

cotton, sugar, wheat, feeder cattle, lean hogs and live cattle. These roll yields are deseasonalized using a filtering procedure based on a locally weighted regression. For a detailed description see Cleveland et al. (1990). The procedure decomposes time series into trend (T), seasonal (S) and remainder components (R):

$$Y_t = T_t + S_t + R_t \tag{2}$$

Our deseasonalized roll yields have the seasonal component subtracted or, in other words it is only the sum of trend and remainder components $(T + R)$. To estimate the seasonal part without forward-looking bias we use an expanding window with the starting window of 10 years (1991-2000). Such a starting window was chosen because some of the commodity roll yield data is available only from 1997. Seasonal effects are also present in the standard deviation of the roll yields. For example, heating oil roll yields in February vary much more than in August. Therefore, buckets based on roll yield are constructed by dividing the deseasonalized roll yield by its' standard deviation of the corresponding month.

We also make use of a second variable based on roll yields which does not suffer from seasonality by construction. It is the monthly roll yield deviation from the 5-year average roll yield of the corresponding month. For example, if this January roll yield is 3% and for the past 5 years roll yield in January on average was 2.5%, then the roll yield deviation is 0.5%. A sector neutral portfolio is constructed using buckets based on such roll yield deviations.

Term structure based investing and roll yields are analyzed in detail by Durr and Voegeli (2009) and Fuertes et al. (2010b). They find that strategies based on term structure significantly outperform the long-only benchmark.

Storage (shipping) costs

There was an idea to use Baltic Dry index growth as an indicator of shipping costs. This index tracks the worldwide international shipping prices of dry bulk cargoes. Since it is not investable, one idea was to construct a portfolio based on the sensitivity to Baltic Dry index. However, it appeared there is not much variation in the sensitivity across different commodities and also the sensitivity varies too much across time.

We have not come across any literature that uses this variable but literature on storage costs dates back to Kaldor (1939) and Working (1949).

Hedging pressure

There are two similar factors used. First one is the hedgers' hedging pressure. It is a sector neutral portfolio long-short in buckets based on net commercial open interest divided by total commercial open interest.

Similarly, speculators' hedging pressure is a sector neutral portfolio long-short in buckets based on net non-commercial open interest divided by total non-commercial open interest.

Open interest is the number of contracts outstanding for a future. The US Commodity Futures Trading Commission (CFTC) produces Commitment of Traders Report with open interest numbers for a variety of commodities. CFTC classifies traders as commercial if they are "commercially engaged in business activities hedged by use of the futures or option markets" as indicated in CFTC Form 40. Therefore, commercial traders are assumed to be hedgers and non-commercial traders to be speculators. For research on these factors see Chang (1985), Bessembinder (1992), De Roon et al. (2000) and Basu and Miffre (2011). All of them find risk premium associated with hedging pressure for futures contracts. Out of the 27 commodities in this research, open interest data is available for 20 of them. It is not available for Brent Crude and Gas Oil which are traded on the Intercontinental Exchange as well as Aluminium, Lead, Nickel, Tin and Zinc on London Metal Exchange. Also, the open interest data starts in 1995 on Datastream but data since 1991 was manually added for several commodities from the CFTC website. The data is adjusted for publication lag as the weekly data for Tuesday's open interest is made available on Friday.

Open interest growth

It is constructed as a sector neutral portfolio long-short in buckets based on the euro value total open interest growth. Since the spot price is not available for the whole timeframe and all commodities, 3-month futures price is used to calculate dollar value open interest. Hong and Yogo (2011) use open interest growth to predict commodity prices. They offer the explanation that open interest can forecast future economic activity and inflation.

Liquidity

We use a sector neutral portfolio long-short according to z-scores (demeaned and divided by the standard deviation) based on absolute returns divided by dollar value total open interest. This measure is based on research by Amihud (2002) on stocks. Since volume data for commodity futures is not available, open interest is used as a proxy.

Market

DJ-UBS Commodity Total Return Index euro returns.

Sector

DJ-UBS Energy Total Return Sub-Index euro returns.
DJ-UBS Industrial Metals Total Return Sub-Index euro returns.

DJ-UBS Precious Metals Total Return Sub-Index euro returns.
DJ-UBS Agriculture Total Return Sub-Index euro returns.
DJ-UBS Livestock Total Return Sub-Index euro returns.

Momentum

Sector neutral portfolios based on the cross sectional z-scores of the past 1, 3, 6 and 12-month returns are used. Pirrong (2005); Miffre and Rallis (2007) and Shen et al. (2007) provide a good research on momentum effects.

Equity

MSCI world index euro returns.
Difference between developed markets and emerging markets equity MSCI indices euro returns.
Difference between MSCI energy and world equity indices euro returns.
Difference between MSCI materials and world equity indices euro returns.
MSCI energy and materials indices start only in January, 1995. The broad MSCI world index represents the market factor for stocks similarly as all non-financial stocks on the CRSP database do in Fama and French (1992). There is no literature that uses other variables (three equity spread returns) in commodity investing.

Interest rates

10 year German bond yield.
Term spread: 10 year minus 2 year German bond yield adjusted to be duration neutral.
Credit spread: Barclays US Corporate Investment Grade index minus Barclays Intermediate US Treasuries index adjusted to be duration neutral.
Default spread: Barclays US High Yield Very Liquid index minus Barclays US Corporate Investment Grade index adjusted to be duration neutral.
Emerging markets spread: JP Morgan Emerging Market Bond Index (EMBI) Global Diversified minus Barclays Long Term US Treasuries index adjusted to be duration neutral.

Spreads have to be duration neutral so that the two legs of the spread react the same to equal yield shifts. Therefore, a spread between two bond indices returns is calculated as follows:

$$Spread_{B_1,B_2} = \frac{R_{B_1} - r_{11}}{D_1} - \frac{R_{B_2} - r_{22}}{D_2}$$

where R_{B_i} is the return of bond i, $i = 1, 2$, r_i is the corresponding risk-free rate and D_i duration.

Variables representing similar factors have been used by Fama and French (1987); Bessembinder H. and Chan (1992) and Vrugt et al. (2007). The emerging markets spread is a new addition to the existing literature. Bond indices instead of specific rating bonds are used because specific bonds exhibit periods of illiquidity and may not be representative at every point in time.

Currency returns

The first variable in this category is the so called EURisk: equally weighted EURUSD, EURGBP, EURJPY, EURAUD returns. This is already used as a factor by ING IM in other asset classes.

Then there are the commodity currencies: EURAUD, EURCAD, EURNZD, EURCLP, EURZAR, EURNOK returns. Since they are heavily correlated, the return of a basket of commodity currencies weighted according to the first principal component is calculated. The first principal component explains 51% of variance. The second and third components would add 18% and 14% of explained variance respectively. The loadings of the first principal component are reported in Table 2 and it is clear that they are different from equal weighting.

Table 2: First principal component loadings

	AUD	CAD	NZD	CLP	ZAR	NOK
1st PC loadings	0.36	0.37	0.31	0.41	0.68	0.12

First principal component loadings calculated by performing a principal components analysis on Euro exchange rate returns against Australian dollar, Canadian dollar, New Zealand dollar, Chilean peso, South African rand and Norwegian krone

All currency returns are adjusted for carry returns, meaning that the money market rates of both currencies in the cross are taken into account. For example, the return of EURAUD is calculated as follows:

$$Ret_{EURAUDt} = \frac{(\frac{S_t - S_{t-1}}{S_{t-1}} + 1) \cdot (MM_{AUDt-1} + 1)}{MM_{EURt-1} + 1} - 1 \tag{3}$$

with S being the spot price of EURAUD and MM the money market rate. LIBOR is always the preference for the money market rate but if it is not available for the full sample period, then the rate supplied by the corresponding central bank is used for the missing observations. If this still does not cover the full sample, deposit rate supplied by the IMF is used.

Finally, JP Morgan Emerging Local Markets Plus Index (ELMI+) return is used that tracks total returns for local currency denominated money market instruments in the emerging markets.

Commodity currencies data used in this research starts in 1999 while ELMI+ starts in 1994. Chen et al. (2010) researches the relationship between a similar set of commodity currencies and commodities. There is no usage of ELMI+ as a factor for commodity returns in the literature.

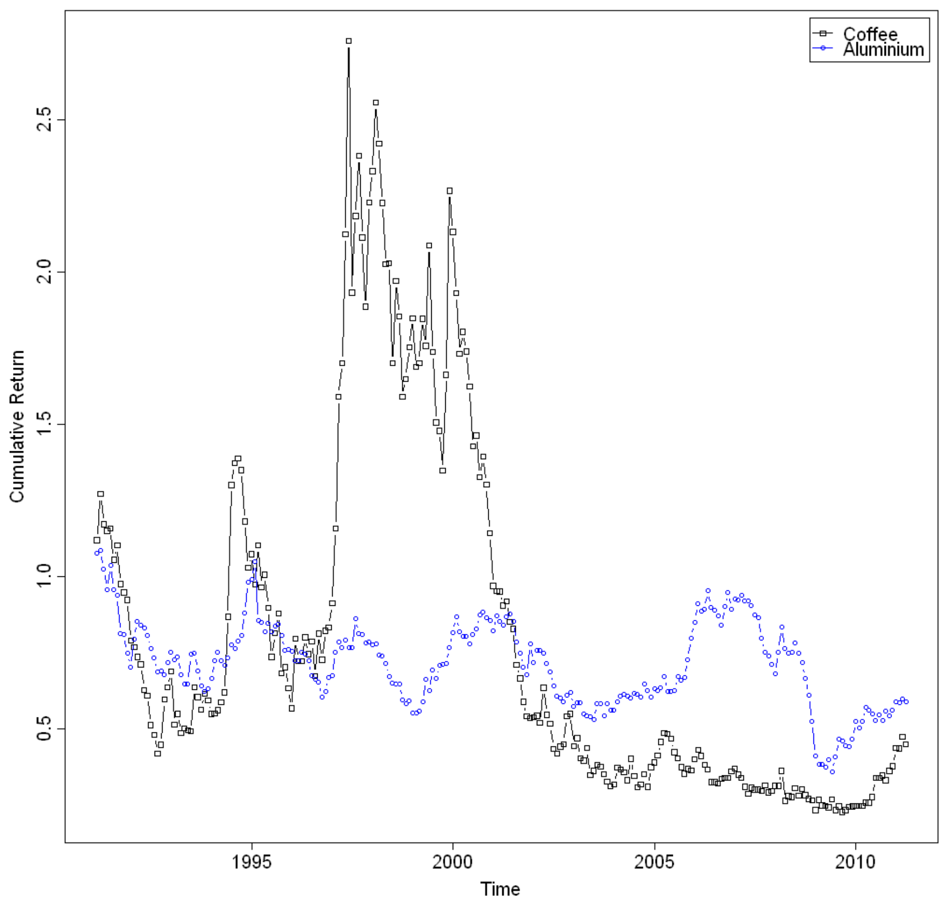
It was considered to include inventories in the dataset, unfortunately there is no reliable, constantly updated data source except London Metal Exchange for a few commodities in the industrial metals sector. Gorton et al. (2007) aggregated the data from many different sources. They find that momentum and roll yield based strategies are partly explained by investing in low inventory commodities, however, they reject the hypothesis that hedging pressure is associated with a risk premium. To sum up, this research deals with 27 commodities in 5 sectors over the time period of 1991 to 2011. There are 34 factors constructed across 9 main themes. In the next section summary statistics of the data are analyzed.

4 Summary statistics

4.1 Commodity returns

Taking the first look at the commodity returns data we see that the performance varies a lot across different commodities. As an example can be coffee and aluminium cumulative returns in Figure 2. During certain time periods there are strong differences in the returns and their variance. Cumulative returns graphs for every commodity are available in Appendix C. There we can notice different returns offered by commodities over our analysis period of 1991-2011.

Figure 2: Coffee and aluminium cumulative returns



To have a better look at this numerically, Table 3 presents summary statistics of several selected commodities from each sector for illustrative purposes. Summary statistics for every commodity can be found in Appendix D.

Table 3: Selected commodities summary statistics

	Brent Crude	Aluminium	Gold	Coffee	Lean Hogs
Min	-0.2723	-0.2181	-0.1155	-0.2998	-0.2616
Median	0.1266	-0.0738	-0.0169	-0.1002	-0.0968
Mean	0.1405	-0.0068	0.0420	0.0401	-0.0789
Max	0.3651	0.1501	0.1568	0.4988	0.2591
StdDev	0.2877	0.1974	0.1572	0.4114	0.2684
Skewness	-0.0399	0.1353	0.3081	0.9675	-0.0522
Kurtosis	4.6209	3.8678	3.7549	5.0671	3.3422

Min and Max are minimum and maximum monthly returns, while Median, Mean and StdDev rows all contain annualized values.

In Table 3 we can clearly observe the difference in the annualized mean returns with lean hogs and aluminium having a negative mean return and brent crude, gold and coffee positive return over our sample. Heterogeneity is present in the annualized standard deviation as well with aluminium and gold being less volatile (0.1974 and 0.1572 respectively) and coffee more volatile with annualized standard deviation of 0.4114. Also, some commodities have positive skewness in returns while others have negative skewness. All commodities exhibit excess kurtosis in comparison to the normal distribution. Every commodity can exhibit very low or very high returns in one month, ranging up to 50% in any direction. From the tables in Appendix D it can be seen that there are differences between commodities from the same sector as well.

4.2 Commodity sectors and aggregate returns

It is also interesting to analyze the performance of the commodity sector indices and the aggregate index over the sample. These indices are used as explanatory factors for individual commodity returns. Table 4 presents the summary statistics.

It seems that some sectors were more profitable than others. Energy, industrial metals and precious metals have positive mean returns while agriculture and livestock mean returns are below zero. As expected from individual commodity summary statistics, sector returns have varying volatility, skew-

Table 4: Aggregate and sector summary statistics

	Aggregate	EN	IM	PM	AG	LS
Min	-0.1323	-0.2425	-0.1943	-0.1216	-0.1992	-0.1709
Median	0.0435	-0.0092	0.0399	0.0476	-0.0236	-0.0431
Mean	0.0366	0.0586	0.0562	0.0589	0.0094	-0.0352
Max	0.1410	0.3545	0.1717	0.1519	0.1299	0.1360
StdDev	0.1513	0.2989	0.2080	0.1755	0.1855	0.1777
Skewness	-0.1620	0.2405	0.0755	0.0709	-0.0530	-0.2163
Kurtosis	3.4037	4.1190	3.3071	3.3446	3.1942	3.1001

Aggregate column is the summary statistics for the DJ-UBS commodity total return index returns. Columns EN, IM, PM, AG and LS are summary statistics for energy, industrial metals, precious metals, agriculture and livestock sector indices returns respectively. Min and Max are minimum and maximum monthly returns, while Median, Mean and StdDev rows all contain annualized values.

ness and kurtosis values. The aggregate index mean returns are on average 0.00305 and have 0.04369 standard deviation with slightly negative skewness.

4.3 Mimicking portfolios returns

There are mimicking portfolios constructed for several factors, namely roll yield, open interest growth, commercial hedging pressure, non-commercial hedging pressure, liquidity and 1, 3, 6 and 12-month momentum. All of them are sector neutral as described in Section 3. Based on the literature overviewed in Section 2, we would expect all of them except non-commercial hedging pressure to have positive mean return because of the construction methodology. For example, commercial hedging pressure portfolio has positive exposure to commodities with high hedging pressure and negative exposure to commodities with low hedging pressure. The literature suggests that hedging pressure is associated with higher risk premium, therefore, such a portfolio is expected to have a positive return. The summary statistics for mimicking portfolios returns are presented in Table 5.

Open interest growth and 1-month momentum portfolios contradict our expectations while the rest is in line with them. Open interest growth factor expectations were formed by the work of Hong and Yogo (2011). However, there are several differences. Firstly, the time period studied is different. Secondly, the set of commodities is different and multiple contracts per commodity are used. Finally and most importantly, they use aggregate commodity market open interest growth to predict aggregate commodity returns. Momentum research expectations were mainly based on Fuertes et al. (2010b) and Fuertes et al. (2010a). The differences in results could have been caused by a different sample period. However, the main deviation might be related to the fact that our momentum portfolios represent the effects of the factor while being sector neutral. If a research does not account for sectoral effects some of the results can be caused by a factor having a heavy exposure to certain sectors.

Since the portfolios are zero-investment, it is best to look at the Sharpe ratio to see which factors are rewarded. It seems that the roll yield and 12-month momentum are the most promising. Profitable strategies and rewarded factors are explored further in Section 8. The rest of the factors summary statistics are of less interest and are available in Appendix E.

Finally, Augmented Dickey-Fuller (ADF) test with 1 lag (selected by Bayesian Information Criterion) was performed both for mimicking portfolio returns and the remaining factor returns. In all of the cases the unit root was rejected at a significance level of 5%. The test statistic values are presented in the last row of Table 5.

Table 5: Mimicking portfolios summary statistics

	Roll Yield	Roll Dev.	OI Growth	CHP	NCHP	Liquidity	Mom1	Mom3	Mom6	Mom12
Min	-0.0536	-0.0878	-0.0721	-0.0520	-0.0779	-0.0705	-0.0406	-0.0633	-0.0567	-0.0481
Median	0.0348	0.0266	0.0008	0.0183	-0.0476	-0.0203	0.0157	0.0271	0.0032	0.0249
Mean	0.0510	0.0145	-0.0135	0.0099	-0.0315	0.0159	-0.0016	0.0248	0.0154	0.0219
Max	0.0770	0.0611	0.1064	0.0459	0.0570	0.0977	0.0614	0.0654	0.0491	0.0669
StdDev	0.0772	0.0829	0.1017	0.0621	0.0682	0.0981	0.0590	0.0651	0.0635	0.0599
Skewness	-0.0990	-0.2530	0.0797	-0.2658	-0.0334	0.4495	-0.0043	-0.1062	0.0020	0.1328
Kurtosis	3.4880	3.7710	3.3716	3.2426	3.8984	3.3656	3.2221	3.9598	3.1531	3.9073
ADF test	-7.5436	-10.1232	-11.4144	-11.5968	-11.1663	-10.9996	-11.3570	-9.9390	-10.5763	-10.4105
Sharpe Ratio	0.6604	0.1749	-0.1328	0.1599	-0.4616	0.1617	-0.0270	0.3811	0.2427	0.3656

The first column is the portfolio based on the roll yield and the second one the portfolio based on the deviations from the 5-year average roll yield. OI Growth is open interest growth portfolio, CHP and NCHP are commercial and non-commercial hedging pressure portfolios respectively. The last four columns are momentum portfolios with 1, 3, 6 and 12-month horizons. Min and Max are minimum and maximum monthly returns, while Median, Mean and StdDev rows all contain annualized values. ADF test row is the ADF test statistic. The ADF 5% significance critical value is -1.95, so a unit root is rejected for all series. The last row contains annualized Sharpe ratios for the portfolios.

5 Univariate analysis

5.1 Methodology

This section presents a framework for the univariate analysis which in later sections leads to the variable selection procedure for the multivariate model. Both time-series and cross-sectional regression models were estimated and compared. Cross-sectional regression is expected to perform well if there are linear relationships between the factors and the returns. However, it cannot capture a non-linear dependency as well as the time-series model. First the time-series analysis is described and later in the section the cross-sectional methodology is also presented.

5.1.1 Time-series

In time-series analysis, for every commodity return we construct a linear regression on every factor as described in Section 3, so 27 regressions on each factor. The exception is the commodity sector factors where only the returns of the corresponding sector's commodities are used, making it 6 for energy, 6 for industrial metals, 3 for precious metals, 9 for agriculture and 3 for livestock. Formally, the regression is

$$R_i = a + bF_j + \varepsilon_{i,j} \tag{4}$$

where R_i is the commodity return, F_j is the factor return, a is the intercept, b is the exposure to the factor and $\varepsilon_{i,j}$ is the error term. A rolling window of 36 months is used. The out-of-sample coefficient of determination (R^2) is calculated meaning that the residual return of this month is computed using the estimation of the previous 36 observations. The mean of the coefficient of determination is reported for every factor. For a factor to be considered in multivariate analysis, the mean R^2 has to be positive as negative R^2 implies unpredictability (see Clark and West, 2006).

It is important that the forecasted exposures match the realized exposures. To test this, we construct characteristic portfolios as described in Grinold and Kahn (2000, chap. 2). These portfolios are zero-investment, minimum variance portfolios that have an exposure of one to a certain factor. Then we test if the returns of these portfolios do not deviate too much from the ex-post actual factor returns. Formally the characteristic portfolio construction can be written down as:

$$\min \, \boldsymbol{h^T V h} \tag{5}$$

s.t.

$$\begin{cases} \boldsymbol{h^T b} = 1, \\ \boldsymbol{h^T e} = 0; \end{cases} \tag{6}$$

where $\boldsymbol{h}$ is the weights vector of the characteristic portfolio, $\boldsymbol{V}$ is the diagonal matrix of the variances of the residuals from the regressions (4), $\boldsymbol{b}$ is the exposures vector of the commodities to the factor and $\boldsymbol{e}$ is a vector of ones. This problem can be solved by formulating the Lagrangian

$$\Lambda(\boldsymbol{h}, \lambda_1, \lambda_2) = \frac{\boldsymbol{h^T V h}}{2} - \lambda_1(\boldsymbol{h^T b} - 1) - \lambda_2(\boldsymbol{h^T e}) \tag{7}$$

and solving the following system of first order conditions:

$$\begin{cases} \boldsymbol{Vh} - \lambda_1 \boldsymbol{b} - \lambda_2 \boldsymbol{e} = 0, \\ \boldsymbol{h^T b} = 1, \\ \boldsymbol{h^T e} = 0. \end{cases} \tag{8}$$

From (8) we get

$$\lambda_1 = \frac{\boldsymbol{e^T V^{-1} e}}{(\boldsymbol{b^T V^{-1} b})(\boldsymbol{e^T V^{-1} e}) - (\boldsymbol{e^T V^{-1} b})^2}, \tag{9}$$

$$\lambda_2 = \frac{-\boldsymbol{e^T V^{-1} b}}{(\boldsymbol{b^T V^{-1} b})(\boldsymbol{e^T V^{-1} e}) - (\boldsymbol{e^T V^{-1} b})^2}. \tag{10}$$

$$\tag{11}$$

Then the solution to (5) with restrictions (6) is

$$\boldsymbol{h} = \frac{\boldsymbol{V^{-1}}(\boldsymbol{e^T V^{-1} e \cdot b} - \boldsymbol{e^T V^{-1} b \cdot e})}{(\boldsymbol{b^T V^{-1} b})(\boldsymbol{e^T V^{-1} e}) - (\boldsymbol{e^T V^{-1} b})^2}. \tag{12}$$

Again, for characteristic portfolio estimation 36 months moving window is used. We compute the returns of the characteristic portfolio at time t with the weights estimated at time $t-1$. These returns are then regressed on the actual factor returns at time t:

$$\boldsymbol{h}_{t-1}^T \boldsymbol{R}_t^{comm} = \alpha + \beta R_t^{factor} + \varepsilon_t. \tag{13}$$

If β is close to one, this means that the factor can be followed by forming a characteristic portfolio. If a characteristic portfolio beta of a factor is significantly different from one with a significance level of 5%, the factor is no longer considered in multivariate analysis.

However, there is one caveat in this approach. It is known that the OLS estimate of the exposures is unbiased. But we use the exposures estimate from (4) to compute the weights of the characteristic portfolio, in other words $\boldsymbol{h}$ is a function of $\hat{\boldsymbol{b}}$ and then we might have

$$\mathbf{E}(\boldsymbol{h}(\hat{\boldsymbol{b}})^{\boldsymbol{T}}\boldsymbol{b} - \boldsymbol{h}(\hat{\boldsymbol{b}})^{\boldsymbol{T}}\hat{\boldsymbol{b}}) \neq 0, \tag{14}$$

which is the same as

$$\mathbf{E}(\boldsymbol{h}(\hat{\boldsymbol{b}})^{\boldsymbol{T}}\boldsymbol{b}) \neq 1. \tag{15}$$

Estimating optimal portfolio weights has been studied by Mori (2004). He analyzes a few different estimators but concludes that neither of them are admissible. To check if there is a bias in our estimation problem and estimate how big it might be, we perform Monte Carlo simulation for every factor and every $\hat{\boldsymbol{b}}$ estimated. We assume that the standard errors in exposures estimation are uncorrelated and

$$\boldsymbol{b} \sim N(\hat{\boldsymbol{b}}, \boldsymbol{V}_{\boldsymbol{\varepsilon}}),$$

where $\boldsymbol{V}_{\boldsymbol{\varepsilon}}$ is the diagonal matrix of standard errors. We generate 5000 vectors of standard errors

$$\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{V}_{\boldsymbol{\varepsilon}})$$

and use antithetic sampling

$$\boldsymbol{\varepsilon}_{i+5000} = -\boldsymbol{\varepsilon}_i, i = 1, .., 5000,$$

effectively resulting in 10 000 $\boldsymbol{\varepsilon}$. The antithetic variates technique was introduced in Hammersley and Morton (1956). It reduces the variance of the samples resulting in greater accuracy with the same sample number. Also, it makes sure that we do not add any extra bias as the mean of our samples is exactly 0. Then we calculate

$$\boldsymbol{b'} = \hat{\boldsymbol{b}} + \boldsymbol{\varepsilon}$$

and find $\boldsymbol{h(b')}$ for every generated instance of $\boldsymbol{\varepsilon}$. Finally, by taking the average of all $\boldsymbol{h(b')}^{\boldsymbol{T}}\hat{\boldsymbol{b}}$, we get the $\mathbf{E}(\boldsymbol{h}^{\boldsymbol{T}}\boldsymbol{b})$. There is substantial bias found.

For example, in the estimation of the market factor the $\mathbf{E}(\boldsymbol{h}^{\boldsymbol{T}}\boldsymbol{b})$ varies from 0.7496 to 0.9598 moving throughout our sample. The average is 0.9190. Therefore, the characteristic portfolio weights are adjusted accordingly:

$$\boldsymbol{h}_{\boldsymbol{adjusted}} = \frac{\boldsymbol{h}}{\mathbf{E}(\boldsymbol{h}^{\boldsymbol{T}}\boldsymbol{b})} \tag{16}$$

This way

$$\mathbf{E}(\boldsymbol{h}^{\boldsymbol{T}}_{\boldsymbol{adjusted}}\boldsymbol{b}) = 1.$$

Note that the adjusted weights still form a zero-investment portfolio. For simplicity from now on we will refer to $\boldsymbol{h}_{\boldsymbol{adjusted}}$ as $\boldsymbol{h}$.

As an alternative, we perform the same procedure with weekly estimation for those variables that such data is available. The choice between monthly and weekly frequency is discussed in Section 3. The same 156-week (36-month) window is used to estimate the exposures but it is expected for them to be more accurate since there are more than 4 times as many observations. We take every month's last week's estimated exposures as exposures for that month and still use monthly returns to compute R^2 and the characteristic potfolios, therefore the results are directly comparable with the monthly estimation analysis. Wednesday is taken as data point for the weekly observation. For example, in June 2000, instead of taking the exposures estimated on June 30th with 36-month window, we use exposures estimated on June 28th (the last Wednesday) with 156-week window. In both of the cases the same monthly return on the last day of the month is used.

5.1.2 Cross-sectional

Alternatively, cross-sectional model can be used. In this case, we can make use of the following factors: roll yield, roll yield deviation, open interest growth, commercial hedging pressure, non-commercial hedging pressure, liquidity, 1, 3, 6, 12-month momentum, commodity market and commodity sectors. For the market factor every commodity has an exposure of one and for the sectors factor commodities get an exposure of one to their sector and zero to all other sectors. For every observation and each factor a cross-sectional weighted least squares (WLS) regression of the commodities return on the factor is carried out and the factor returns estimated. In the end we have the residual returns series for every commodity. Since in this case the R^2 is not out-of-sample, we calculate and report the adjusted coefficient of determination for every commodity:

$$R^2_{adj} = 1 - \frac{SS_{res}(n-1)}{SS_{tot}(n-p-1)} \tag{17}$$

where SS_{res} and SS_{tot} are the residual and total sum of squares respectively, n is the sample size and p is the number of factors.

In a similar fashion as the test with characteristic portfolio returns in time-series analysis, here we also have an additional criterion besides adjusted R^2. The model assumes uncorrelated residual and factor returns. Our criterion is that the model cannot have a significantly different from zero correlation with a significance level of 5%. To test for this we need to apply the Fisher transformation to the sample correlation coefficient r (see Fisher, 1921):

$$z = arctanh(r) = \frac{1}{2} \ln \frac{1+r}{1-r} \tag{18}$$

where *arctanh(r)* is the inverse hyperbolic tangent. Then z is approximately normally distributed with mean *arctanh*(ρ) and standard deviation $\frac{1}{\sqrt{n-3}}$ where ρ is the population correlation coefficient and n is the sample size. If we find significant correlation between a factor returns and commodity residual returns, we do not use that factor.

5.2 Results

In this section we present the results obtained by implementing the techniques discussed in Section 5.1. First the time-series results are presented and then the cross-sectional analysis follows. At the end of the section we make a choice for the model that is going to be used for the multivariate analysis.

5.2.1 Time-series

In Table 6 the average out-of-sample R^2 of the factors is reported along with the characteristic portfolio beta and the p-value of this beta being significantly different from one.

From Table 6 it can be seen that as expected the highest R^2 is attributed to commodity market and sectors factors. This is not surprising since both market and sector factors are directly composed of individual commodities.

Table 6: Univariate monthly time-series analysis results

	Mean OOS R^2	Char Port Beta	p-value
Roll Yield	-0.0188	0.8806	0.6695
Roll Yield Dev	-0.0316	0.3831	0.0102
EURisk	0.0304	0.2862	0.0037
CC-PC1	0.0068	-0.4715	0.0007
AUD	0.0520	0.3060	0.0391
CAD	0.0205	-0.2512	0.0053
NZD	-0.0096	0.3161	0.0882
CLP	-0.0237	0.1627	0.0022
ZAR	-0.0325	-0.9236	0.0003
NOK	0.0038	0.3147	0.0070
ELMI	0.0007	0.2598	0.0493
MSCI World	0.0350	0.5226	0.0664
EM Spread Eq	-0.0023	0.4165	0.0038
Energy Spread Eq	0.0171	0.9483	0.7095
Materials Spread Eq	0.0148	0.3471	0.0145
Comm Market	0.1843	0.9056	0.3530
Sector Energy	0.6554	1.6191	0.6164
Sector IndMet	0.4999	0.8104	0.0714
Sector PrecMet	0.6163	1.5160	0.0519
Sector Agriculture	0.3186	0.6343	0.1690
Sector Livestock	0.6745	1.0644	0.7894
Long Rate	-0.0139	0.1488	0.0095
Term Spread	-0.0703	-0.1009	0.0000
Credit Spread	-0.0314	0.5182	0.0255
Default Spread	-0.0276	0.7591	0.3160
Emerging Spread	0.0093	0.7671	0.5338
Open Interest Growth	-0.0321	0.3285	0.0442
Hedge Comm	-0.0240	0.5827	0.0704
Hedge NonComm	-0.0392	0.1926	0.0005
Liquidity	-0.0149	0.8532	0.1165
Mom1	-0.0367	0.0745	0.0001
Mom3	-0.0421	0.3713	0.0048
Mom6	-0.0322	0.3076	0.0005
Mom12	-0.0378	0.1970	0.0002

Note that in this case some of the R^2 is gained because the same component is on both sides of the regression. However, we can show a rough estimation that this does not change our decision. The maximum gain in "free" R^2 is when the sectors are combinations of equally weighted commodities with equal variances. In that case the market factor maximum gain is $(\frac{1}{27})^2 = 0.0014$ while sector factors maximum gains are:

- $(\frac{1}{6})^2 = 0.0278$ for Energy
- $(\frac{1}{6})^2 = 0.0278$ for Industrial Metals
- $(\frac{1}{3})^2 = 0.1111$ for Precious Metals
- $(\frac{1}{9})^2 = 0.0123$ for Agriculture
- $(\frac{1}{3})^2 = 0.1111$ for Livestock

Even adjusted for these maximum gains in R^2 the sector factors would have the highest R^2.

Other factors that look promising are the MSCI World return, energy sector equity spread and emerging markets bonds spread as they have positive R^2 and characteristic portfolio betas that are not significantly different from one with the significance level of 5%. Also, some currency returns and equity materials sector spread have positive R^2 but their characteristic portfolio betas differ from one significantly.

As described in Section 5.1.1, the same analysis is carried out with weekly estimation but still using monthly returns together with the last week's exposures estimates each month. The results of this univariate analysis are presented in Table 7. The roll yield and roll yield deviation factors are not available as they are based on monthly seasonality. Also, take note that momentum factors are now 4, 13, 26 and 52-week to be in line with the monthly momentum factors. It can be observed in Table 7 that the results improved in the sense that most of the factors have positive and higher R^2. Also, the hypothesis of the characteristic potfolio beta being equal to one cannot be rejected with significance level of 5% in most of the cases. Again, unsurprisingly, commodity market and sectors factors have the highest R^2. However, the market factor average out-of-sample R^2 of 0.23 is lower than of any sector. Note that again for market and sector factors some of the R^2 is gained because the same component is on both sides of the regression. However, the rough estimation showed previously that even adjusted for this, it does not change the fact that sector factors have the highest R^2. Using weekly data, bond term spread, open interest growth and 4-week momentum factors can be immediately rejected because of negative R^2 which means these factors have no predictive ability. In addition, MSCI World return, emerging equity spread, energy sector equity spread and commercial hedging pressure as well as 26 and 52-week momentum factors do not pass the characteristic portfolio test with betas being different from one. Based on these results weekly time-series is the preferred method over the monthly time-series. In the multivariate analysis only the factors that have a positive mean R^2 and the characteristic portfolio beta test p-value is above 0.05 will be considered.

Table 7: Univariate weekly time-series analysis results

	Mean R^2	Char Port Beta	p-value
EURisk	0.0933	1.7028	0.0685
CC-PC1	0.0472	0.9092	0.8279
AUD	0.0682	1.8144	0.1010
CAD	0.0609	0.6309	0.5269
NZD	0.0288	1.8353	0.1579
CLP	0.0215	0.9259	0.8491
ZAR	0.0043	1.8234	0.1057
NOK	0.0485	0.4534	0.2013
ELMI	0.0405	1.7718	0.0597
MSCI World	0.0699	1.7959	0.0354
EM Spread Eq	0.0301	2.7570	0.0101
Energy Spread Eq	0.0515	2.0910	0.0128
Materials Spread Eq	0.0444	1.0874	0.8743
Comm Market	0.2257	1.1726	0.1085
Sector Energy	0.7130	0.7703	0.2778
Sector IndMet	0.5195	0.9383	0.5664
Sector PrecMet	0.6298	0.6774	0.3327
Sector Agriculture	0.3375	0.8774	0.4084
Sector Livestock	0.6968	1.0150	0.9402
Long Rate	0.0271	1.8416	0.1209
Term Spread	-0.0070	2.0442	0.0666
Credit Spread	0.0326	1.2850	0.5932
Default Spread	0.0275	1.4590	0.1843
Emerging Spread	0.0383	0.9422	0.9320
Open Interest Growth	-0.0062	-0.1947	0.0387
Hedge Comm	0.0175	1.5938	0.0254
Hedge NonComm	0.0075	1.1286	0.6201
Liquidity	0.0269	1.0325	0.6423
Mom4	0.0057	1.2421	0.5054
Mom13	0.0079	1.1910	0.4738
Mom26	0.0006	1.2192	0.4310
Mom52	-0.0040	1.0171	0.9444

5.2.2 Cross-sectional

There is also cross-sectional analysis with monthly frequency carried out. In Table 8 the results following methodology as described in Section 5.1.2 are presented.

Table 8: Univariate monthly cross-sectional analysis results

	Mean R^2	Failed comm.	Total comm.
Roll Yield	-0.0128	13	27
Roll Yield Dev	0.0349	5	27
Comm Market	0.2119	7	27
Sector Energy	0.7434	2	6
Sector IndMet	0.4417	2	6
Sector PrecMet	0.3908	1	3
Sector Agriculture	0.2573	3	9
Sector Livestock	0.5040	0	3
Open Interest Growth	0.0173	0	27
Hedge Comm	0.0056	0	27
Hedge NonComm	-0.0113	14	27
Liquidity	0.0083	1	27
Mom1	0.0321	0	27
Mom3	0.0301	12	27
Mom6	0.0406	4	27
Mom12	0.0319	6	27

The first column is the average adjusted R^2. The second one contains the number of commodities that did not pass the residual and factor returns correlation test and the third one the total number of commodities for the corresponding factor.

The first column in Table 8 lists the mean adjusted R^2 for every univariate WLS regression. The second column represents the number of commodities that failed the residual returns correlation test, meaning that their residual returns correlation with factor returns was significantly different from zero with the significance level of 5%. It can be observed that the only factors that have positive mean adjusted R^2 and all commodities passing the test are the livestock sector, open interest growth, commercial hedging pressure and 1-month momentum. Therefore, our preferred model choice remains the weekly estimation time-series model.

6 Multivariate analysis

This section builds on the previous one by taking the chosen weekly estimation time-series model and extending it with more factors following customized stepwise regression procedure with several factor inclusion criteria. As in the previous section, first the methodology is described and later on the multivariate results are presented and the final factor model is stated.

6.1 Methodology

The model construction is based on our own customized stepwise regression procedure. There are three criteria that are taken into account when adding extra factors: the mean out-of-sample R^2, the characteristic portfolio beta test p-values and the condition number. The starting pool of factors include only those that had a positive mean R^2 and characteristic portfolio beta not significantly (5% significance level) different from one in the weekly univariate analysis. First, we start with the factor that has the highest R^2 and is managable according to the characteristic portfolio test. Then we proceed adding factors with the highest R^2 that do not violate one of the following criteria. One is that all the factors in the model have to managable (characteristic portfolio beta test passed). If at least one of them is not, the model is rejected. The other criterion is that the mean out-of-sample R^2 should not decrease when adding extra factors. Finally, the correlation matrix of the commodity factors together with the factors that ING IM uses for other asset classes has to be sufficiently invertible in order to be able to estimate portfolio weights later on. The list of current factors in other asset classes is available in Appendix F. We measure the invertibility of the correlation matrix by looking at its' condition number. A larger matrix condition number means that changes in the argument matrix lead to bigger changes in the result than with a smaller condition number. The matrix $\boldsymbol{A}$ condition number $\kappa(\boldsymbol{A})$ is calculated as the product of two norms:

$$\kappa(\boldsymbol{A}) = \|\boldsymbol{A}\| \cdot \|\boldsymbol{A}^{-1}\| \tag{19}$$

In this case with the Euclidian norm it is

$$\kappa(\boldsymbol{A}) = \frac{\sigma_{max}(\boldsymbol{A})}{\sigma_{min}(\boldsymbol{A})} \tag{20}$$

where σ is a singular value. Since correlation matrix is symmetric, we get

$$\kappa(\boldsymbol{A}) = \frac{\lambda_{max}(\boldsymbol{A})}{\lambda_{min}(\boldsymbol{A})} \tag{21}$$

where λ represents an eigenvalue of matrix $\boldsymbol{A}$. We also require all eigenvalues of the correlation matrix to be positive. It is desired that added factors do not inflate the correlation matrix condition number or else the inverse will be less accurate. If an extra factor increases the condition number by more than 5%, the model is rejected and the factor is excluded from the subsequent steps as well.

Following these three criteria, factors are included in the model one-by-one. The final model is the one that passes all criteria and has the maximum R^2.

6.2 Results

It is obvious from the univariate results that our starting model has five commodity sectors returns as the first factors. Every commodity is assigned only one sector as explanatory factor and the characteristic portfolios are constructed by using commodities in the corresponding sector only. Such a model has the mean out-of-sample R^2 of 0.5338, condition number of 153.1147 and all five factors pass the characteristic portfolio beta test. From here we follow the customized stepwise regression procedure described in 6.1 to arrive at the final model of our choice. Almost 30 models with a different set of factors were tested in total during the stepwise procedure. The best models of each step that passed the procedure (including the final one) are presented in Table 9.

Table 9: Best models leading to the final choice

Model no.	R^2	C. no.	p1	p2	p3	p4	p5	p6	p7	p8	p9
I	0.5338	153.1147	0.2778	0.5664	0.3327	0.4084	0.9402				
II	0.5573	156.0369	0.9642	0.2480	0.3568	0.2359	0.6630	0.2461			
III	0.5652	156.0369	0.9943	0.2164	0.4564	0.2064	0.5766	0.2463	0.1350		
IV	0.5711	158.5184	0.8708	0.2156	0.4509	0.2181	0.5647	0.2086	0.0948	0.6465	
V (final)	0.5740	165.0 084	0.7936	0.2098	0.4821	0.2014	0.6250	0.6370	0.1090	0.0958	0.8569

The first column is the model number. Each model has one more factor than the previous one. I: Sectors; II: Sectors and liquidity; III: Sectors, liquidity and materials equity; IV: Sectors, liquidity, materials equity and 3-month momentum; V: Sectors, liquidity, materials equity, 3-month momentum and non-commercial hedging pressure. The second column lists the mean out-of-sample R^2. The third column contains the condition number. Columns 4-12 contain the p-values of the characteristic portfolio beta test for every factor. p1-p5 are the p-values for the sector factors (energy, industrial metals, precious metals, agriculture, livestock), p6-p9 are the p-values for the liquidity, materials equity, 3-month momentum, and the non-commercial hedging pressure factors, respectively.

In the second step the best factor to expand the model was liquidity. It increased the mean out-of-sample R^2 to 0.5573 without failing other citeria outlined in the stepwise regression procedure. The third model had materials sector equity factor added and R^2 of 0.5652. Note that the condition number did not change because materials sector equity is already in the ING IM factor list in Appendix F. The next step included 3-month commodity momentum and increased the R^2 to 0.5711. The final model consists of five sectors, liquidity, materials equity, 3-month momentum and non-commercial hedging pressure factors. It has the mean out-of-sample R^2 of 0.5740 and the condition number of 165.0084. Note that this can be adjusted for the R^2 gained by having sector factors that include commodity returns themselves. Following the rough calculation in 5.2.1, we can estimate the weighted average maximum gain of "free" R^2 from the sector factors to be 0.0412. This would leave the final model R^2 at 0.5328. The sector factors represent the exposure to the systemic risk of every sector. Materials equity sector factor is similar in this sense. It is directly related to commodity prices as the firms in the materials sector deal with discovering and refining raw materials. The liquidity factor should represent the premium that is paid for holding less liquid contracts. Momentum can stem from both behavioural bias and, as Gorton et al. (2007) suggests, from bearing risk when inventories are low. The non-commercial hedging pressure factor relates to the risk transfer from the hedgers to the speculators. However, after inspecting the factors in more detail, we raise concerns on what the liquidity factor actually represents. The average cross-sectional liquidity ranks for every commodity are presented in Table 10.

From the average ranks we conclude that the measure regards Copper, Corn and Gold as the most liquid while on the other hand Sugar, Natural Gas and Silver as the least liquid of the commodities that we had data for. Keep in mind that the factors are formed using sector neutral buckets. This means that almost always the liquidity factor is long in silver (the illiquid precious metal) and short in gold (the liquid precious metal). However, by looking at this data, we were not convinced that it is indeed liquidity that is captured here. Some further research on this is recommended. The volume data could be useful which we did not have.

As an example of the stepwise regression procedure, in Tables 11-13 there are presented a few models that failed one or more of the procedure criteria. From Table 11 it can be observed that the emerging markets bonds factor was rejected because the R^2 decreased to 0.5277 from 0.5338 in model I (Table 9).

Table 10: Liquidity average ranks

Commodity	Mean rank
Crude Light	6.97
Heating Oil	7.07
Natural Gas	12.11
Copper	2.58
Gold	2.85
Silver	12.27
Cocoa	3.24
Coffee	9.86
Corn	2.73
Cotton	9.73
Sugar	11.67
Wheat	4.00
Feeder Cattle	10.19
Lean Hogs	10.66
Live Cattle	6.72

The second columns reports the average rank of the commodity according to our liquidity variable. Some of the commodities are missing in this table because we did not have the data to calculate the liquidity measure.

Also, the condition number increased too much. Model VII was unmanagable in terms of characteristic portfolio beta test with three p-values below 5% and the condition number got very high because the market and sector factors overlap too much. In the final example in Table 13 credit spread factor did increase the R^2 more than the materials sector equity returns in model III (0.5667 vs 0.5652 in Table 9), however, it also increased the condition number beyond our threshold to 174.6081. By following such step-wise regression procedure factors kept falling off and we have finally arrived at model V presented previously in Table 9 with the R^2 of 0.5740.

Table 11: Rejected model example 1

Model no.	R^2	C. no.	p1	p2	p3	p4	p5	p6
VI	0.5277	172.0604	0.9742	0.4054	0.4524	0.3097	0.2880	0.1374

Model VI consists of five sectors and the emerging markets bonds spread. The second column lists the mean out-of-sample R^2. The third column contains the condition number. p1-p5 are the p-values for the sector factors (energy, industrial metals, precious metals, agriculture, livestock), p6 is the p-value for the emerging markets bonds factor.

Table 12: Rejected model example 2

Model no.	R^2	C. no.	p1	p2	p3	p4	p5	p6
VII	0.5418	1876.6550	0.0238	0.1094	0.0065	0.0695	0.0030	0.2568

Model VII consists of five sectors and the aggregate commodity market return. The second column lists the mean out-of-sample R^2. The third column contains the condition number. p1-p5 are the p-values for the sector factors (energy, industrial metals, precious metals, agriculture, livestock), p6 is the p-value for the aggregate commodity market factor.

Table 13: Rejected model example 3

Model no.	R^2	C. no.	p1	p2	p3	p4	p5	p6	p7
VIII	0.5667	174.6081	0.9764	0.2065	0.3432	0.2035	0.6528	0.2299	0.7641

Model VIII consists of five sectors, liquidity and the credit spread factors. The second column lists the mean out-of-sample R^2. The third column contains the condition number. p1-p5 are the p-values for the sector factors (energy, industrial metals, precious metals, agriculture, livestock), p6 is the p-value for the liquidity factor and p7 is the p-value for the credit spread.

7 Model testing

Having established the final model choice, we perform the Gibbons-Ross-Shanken (GRS) test of efficiency and present the results in this section. The GRS test documented in Gibbons et al. (1989) and applied for multifactor models in Cochrane (2000) tests the null hypothesis that the intercepts of the test assets regressed on the multifactor model jointly equal zero. If the null is rejected it means that the deviations are statistically significant and the factors do not explain all returns. Commonly it is assumed that this is due to missing factors (for example in Fama and French (1993)) but MacKinlay (1995) advocates an alternative view. He also provides a summary of the literature on this issue. The test statistic is computed (see Cochrane, 2000, chap. 12) as follows:

$$W = \left(\frac{T-N-L}{N}\right)\left(\frac{\hat{\alpha}^T\hat{\Sigma}^{-1}\hat{\alpha}}{1+\bar{\mu}^T\hat{\Omega}^{-1}\mu}\right) \tag{22}$$

where T is the number of observations, L is the number of factors and N is the number of test assets. $\hat{\alpha}$ is the vector of estimated intercepts, $\hat{\Sigma}$ is the estimate of the residual covariance matrix, $\bar{\mu}$ is the vector of factor portfolios means and $\hat{\Omega}$ is the unbiased estimate of factor portfolios covariance matrix. It is clear that if all intercepts are equal to zero, then the test statistic is zero indicating that the null cannot be rejected. Gibbons et al. (1989) showed that the test statistic follows the Fisher distribution with N and $T-N-L$ degrees of freedom. However, for the test to follow the Fisher distribution it requires normal, uncorrelated and homoskedastic residuals (see Cochrane, 2000, chap. 12). Also, the GRS test assumes that factors are priced without error.

Our test assets are the 27 commodity indices. The factors are the ones from model V described in Section 6.2: commodity sectors, liquidity, equity materials sector, 3-month momentum and non-commercial hedging pressure. By running the 27 regressions we get the GRS test statistic estimate of 1.6754 that corresponds to the p-value of 0.0275. Therefore, we reject the null hypothesis at 5% significance level meaning that there is at least one commodity that is not fully explained by the model. A few issues can be of concern here. The assumption that the factors are priced without error which can be too restrictive. Secondly, there could be correlation among the residuals. In fact, it is shown in Grinold and Kahn (2000, chap. 3) that the residual correlation of stocks in general must be negative. An alternative test that is not implemented in this thesis but could be considered in further research is the Generalized Method of Moments (GMM) (see Cochrane, 2000, chap. 12). It is more flexible and relies on fewer assumptions. GMM relaxes the assumption that factors are priced without error and does not assume uncorrelated residuals. However, as opposed to the GRS, it is not a finite sample test.

8 Strategies

The research done on the risk model is a good basis for the strategy development. The risk research provides guidance in getting a sense in what drives the commodity returns. Therefore, the insights from the first part of the research help in developing ideas for the strategies. This section describes a list of timing indicators that might prove to be useful in the strategies. Later on in the section the methodology for two different types of strategies is provided. Finally, in the results section the strategies that were tested are described with the economic intuition behind them and the results presented.

The following is a list of timing indicators that are used in construction of the strategies. References are provided if there is any literature of the indicators used in commodity research.

OECD Composite Leading Indicator

We use the OECD Composite Leading Indicator with restored trend for the OECD economies and 6 non-member economies (Brazil, China, India, Indonesia, Russia and South Africa). For the mechanics on the index construction and techniques applied, details can be found in Nardo et al. (2008) and Gyomai and Guidetti (2008). There is a 2-month publication lag which is taken into account when constructing strategy signals. To better measure the direction and the speed of change of the world economy, we take a 6-month Simple Moving Average indicator (the difference between the monthly difference and the 6-month average monthly difference) of CLI instead of monthly CLI difference. Such an indicator has been used by ING IM in equity research and it was suggested to use the same construction. There is no academic literature that makes use of OECD CLI in commodity strategies.

Risk Aversion

Risk Aversion Index (RAI) deviation from its' 3-month average. It is a proprietary ING IM aggregate index that tracks the risk aversion in three sub-groups - Liquidity, Credit and Volatility. RAI consists of 11 indicators, 3 measuring liquidity, 4 measuring credit and 4 for volatility. Liquidity indicators include spread between the benchmark 10-year US Treasury note and the most recent off-the-run 10-year note, US swap spread and the US term spread between 2-year and 3-month yields. Credit sub-group consists of JP Morgan EMBI Index which measures sovereign spread, US corporate high yield spread, MSCI Emerging Markets Free Index and equally weighted MSCI World IT and MSCI World Telecoms indices. Finally, in volatility sub-group there are equally weighted 3-month implied volatilities for EURUSD, CADUSD, AUDUSD and JPYUSD exchange rates, VIX (implied equity volatility from options), implied Brent Crude volatility and implied Gold volatility.

Obviously, this index has never been used in the literature but the relationship between risk appetite and commodity returns is studied by Etula (2009). However, he takes innovations in the US broker-dealer aggregate balance sheets as a proxy for risk appetite.

Speculator activity

The ratio between speculators and hedgers open interest. The speculators and hedgers open interest ratios to the total open interest are reported by the CFTC. Our measure is the ratio of these two ratios. If it is more than one, speculators make up the bigger share of the total open interest and therefore are regarded as more active than the hedgers.

We also make use of momentum, roll yield, hedging pressure and open interest growth factors from Section 3. In the next section the process of developing strategies and testing them is described.

8.1 Methodology and testing

In this section the methodology for constructing the strategies, measuring their performance and then combining the best ones is laid out.

8.1.1 Construction

We aim for a robust strategy development as it is very important to minimize the risk of the performance of the strategies diminishing going forward in time. Therefore, one of the base rules in this research is that every proposed strategy must have an economic rationale behind it before it can be tested. Also, if arbitratry choices have to be made (for example a rolling window size), robustness checks are carried out afterwards. On the other hand, no data mining to select such values for the best performance is done. For all the strategies the research is done on monthly signals and returns.

There are two different ways of how strategies are applied: across sectors and within sectors. The first way is to estimate the differences between sectors according to some signal and go long in two sector indices, short in two and neutral in the fifth one. The other type of strategies are neutral on the sector level but go long $\frac{1}{3}$ and short $\frac{1}{3}$ commodities within each of the sectors. Both of these types are zero-investment. All signals are either -1, 0 or 1, meaning go short, neutral or long with the same bet size across all strategies. Below is a list of strategies proposed with the economic intuition behind them explained. All of the choices that are made either have the reasoning provided or it is explicitly stated that it is an arbitrary choice. Every strategy is tested for both across sectors and within sectors performance. In the across sectors case if a characteristic is not available on the sector level (for example roll yield), then the signal for a sector is based on the average characteristic of individual commodities in that sector (for example the average roll yield of energy commodities).

1-month momentum
: Go long in commodities with high previous month volatility adjusted returns and short in low.

3-month momentum
: Go long in commodities with high previous 3-month volatility adjusted returns and short in low.

6-month momentum
: Go long in commodities with high previous 6-month volatility adjusted returns and short in low.

12-month momentum
: Go long in commodities with high previous 12-month volatility adjusted returns and short in low.

All momentum strategies signals are divided by exponentially weighted volatility estimate

$$Mom_{adj\,t} = \frac{Mom_t}{\sigma^2_{t+1}} \tag{23}$$

with a recursive formula for exponentially weighted variance estimate

$$\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)r_t^2 \tag{24}$$

Here the decay is $\lambda = 0.93$ with estimation window of 36 months. Mom is the pure momentum signal, Mom_{adj} is the volatility adjusted momentum signal and r is the return. This is done in order to prevent the strategy from going into more volatile commodities most of the time. Such a volatility estimation has been chosen by ING IM and is used in other asset classes as well. Momentum strategies are tested because they seemed to be rewarded in Table 5 and also they have been backed by the literature overviewed in Sections 2 and 3.

Roll yield

Go long in commodities with high roll yield and short in low. The roll yield is calculated in the same way as in Section 3 except that no deseasonalizing procedure is applied. Also, instead of taking only the roll yield between 1-month and 3-month futures, here we use the sum of the roll yields between 1-month and 3-month futures and between 3-month and 6-month futures. By using (1) and applying it for the current case, we get

$$\begin{aligned}\text{Roll yield} &= \frac{f_1 - f_3}{2} + \frac{f_3 - f_6}{3} + 2r \\ &= \frac{3f_1 - f_3 - 2f_6}{6} + 2r \end{aligned} \tag{25}$$

The reason for using the sum of the roll yields is that not only high near-term roll yield matters but also how stable the futures curve structure is. If there is high near-term roll yield but the slope of the futures curve changes with maturity longer than 3 months, it can be an indication of instability.

Roll yield 5-year deviation

Go long in commodities with large positive deviation from the 5-year average of the corresponding month's roll yield and short in large negative deviation. Calculated in the same way as in Section 3 except that the sum of two roll yield deviations is used for the reasons of stability, just as above. Roll yield strategies are motivated by the results in Table 5 and positive results reported by other studies as described in Sections 2 and 3.

Hedgers' hedging pressure

Go against the positions of hedgers. Long where hedgers are mostly short and vice versa, same way as described in Section 3. This is also one of the wider described theories and is discussed in Sections 2.

Risk sensitivity

This strategy assumes that investors perceive some commodities as more risky and others as less risky (safe haven). We measure the rolling correlation of each commodity return to the deviation of Risk Aversion Index from its' 3-month average. This way we try to estimate which commodities currently are perceived to be relatively safe. If the RAI deviation is positive now (risk aversion is growing), we go long in the most positively correlated commodities (safe-haven) and short in the most negatively correlated (risky). If the RAI deviation is negative, we do the opposite. The window used for rolling correlation estimation is 2 years with weekly frequency, although the strategy signals are still monthly as for any other strategy. The choice of 2-year rolling correlation window was arbitrary to avoid any optimization that may lead to instable future performance. This strategy was partly motivated by the perception of gold as a safe haven commodity, especially during the last 5-10 years. Gold as a hedge or safe haven against equities or bonds is analyzed by Baur and Lucey (2010); Baur and McDermott (2010). They find that gold is a safe haven against developed markets equity but not against bonds or emerging markets equity. Our risk sensitivity strategy is set up in a similar fashion but it is dynamic and broader.

Economic cycle sensitivity

Very similar to the previous strategy except that instead of risk, the economomic cycle sensitivity is measured. The rolling correlation of commodity returns with the 6-month Simple Moving Average (SMA) rule of CLI is measured. If the SMA of CLI is positive (economy is expanding), long positions are taken in commodities with the highest positive correlation (cyclical) and short in the highest negative correlation (defensive) and vice-versa if the SMA of CLI is negative. The window used for rolling correlation estimation is 2 years (monthly frequency as OECD CLI is published monthly) and was chosen arbitrarily. I have not come across any literature with a similar strategy but in equities it is common to group equity sectors into "cyclical" and "defensive" (see Froot and Teo, 2008).

12-month momentum with open interest growth confirmation

This is an enhanced momentum strategy that uses open interest growth to confirm the trend. If the open interest growth is positive, then the signal for a certain commodity is the same as for 12-month momentum strategy. However, if the open interest growth is negative, the signal is set to zero. The logic is that open interest growth indicates strengthening (weakening) of the support for the current price with the market being more (less) active. We have not come across any academic papers that analyze a similar strategy but it is described by Murphy (1999, chap. 7). This type of strategy is also supported by the findings of De Grauwe and Grimaldi (2006) who conclude that self-fullfilment feature of technical strategies is part of their success.

Speculators' hedging pressure

Go along the positions of speculators. Long where speculators are mostly long and short where they are mostly short, same way as described in Section 3. Chang (1985) and Basu and Miffre (2011) find support for this theme.

Overspeculation reversal

A strategy that tries to predict a reversal in speculators' positions. The idea is that if the futures of a commodity are traded mainly by speculators, it should revert back to a price where it is supported by more equal activity of hedgers and speculators. Here we use the speculator activity variable described earlier. If it is above its' 80th percentile during the past 2 years (weekly frequency), we take positions against the speculators. The signal then is the opposite to speculators' hedging pressure signal. The signals and returns are still on monthly basis. The 2-year window and the 80th percentile were arbitrary choices. There is no academic research on such a strategy.

8.1.2 Performance measures

After we have established a list of strategies, each of them is tested according to the same framework. All strategies are scaled to have 1% historical volatility in order to have a direct comparison of cumulative returns and geometric Information Ratio. There are several measures of performance. The first one is the information ratio (IR). It is calculated as $IR = \frac{\alpha}{\sigma(\varepsilon)}$ where α is the Jensen's alpha and ε is the residual from regressing the strategy returns on the benchmark. Dow Jones-UBS Commodity Index return is the benchmark in this case. A slightly modified version is the geometric IR (GIR). It is computed in a similar manner as IR but with geometric mean and geometric standard deviation instead of arithmetic ones:

$$GIR = \frac{GM(\alpha)}{GSD(\alpha)} \tag{26}$$

The disadvantage of IR is that it does not consider compounded returns, therefore, one might end up with positive IR but a negative cumulative alpha.

To test the strategy performance the non-parametric Henriksson-Merton (HM) timing test is carried out. It measures the directional performance of the strategy. The test says that a strategy is successful (rejects the null hypothesis of no market timing ability) if the sum of the conditional probabilities of a correct forecast is larger than one. A correct forecast is defined as a long signal when the market is above the median (up) or a short signal when the market is below the median (down). The HM test statistic is calculated as follows:

$$HM = \frac{n_{11} - \frac{n_{10}n_{01}}{n}}{\sqrt{\frac{n_{10}n_{01}n_{20}n_{02}}{n^2(n-1)}}} \tag{27}$$

where:

- n_{11} is the number of correct up market forecasts (meaning the signal is long and the market is above the median)
- n_{01} is the number of up markets
- n_{02} is the number of down markets
- n_{10} is the number of long signals
- n_{20} is the number of short signals
- n is the number of observations

The test statistic is asymptotically standard normally distributed under the null hypothesis. One can refer to Henriksson and Merton (1981) or Pesaran and Timmermann (1994) for a detailed description of the test. Also, there is an application of the HM test in Marquering and Verbeek (2004). In addition, we apply Bayesian inference (see, for example, Box and Tiao (1973)) to estimate the expected information coefficient (IC) and the probability that it is above 0.01. From the Bayes' theorem we have

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)} \tag{28}$$

where Y is the observed data (the number of correct forecasts) and it follows the binomial distribution. θ is the unknown parameter which is the probability that the forecast is correct. By using a flat prior $P(\theta)$ and calculating the marginal probability mass function $P(Y) = \sum_{\theta} P(Y|\theta)P(\theta)$ we estimate the posterior $P(\theta|Y)$. In the case of market timing where there are only two outcomes of the market (above median or below), the probability of a correct forecast translates into the IC, assuming we want to only forecast the direction (see Grinold and Kahn, 2000, chap. 6). Then the posterior mean $\mathbf{E}[P(\theta|Y)]$ is the expected IC. $\mathbf{E}[P(\theta > 0.01|Y)]$ and $\mathbf{E}[P(\theta < -0.01|Y)]$ are the expected probabilities that the IC is above 0.01 or below -0.01 respectively. There are also several other statistics calculated, like skewness, kurtosis or the beta to the benchmark.

8.1.3 Combination

There is a two-step process establishing the best strategy combination. In the first step, the geometric Information Ratio has to be positive and the HM test p-value below 0.3. The 0.3 p-value threshold was chosen by ING IM because the cost of not rejecting the null hypothesis when the alternative is true (type II error) and thus missing out on a profitable strategy is relatively high to the cost of rejecting the null when the null is true (type I error) and thus having a strategy that contains only noise. Therefore, it is preferred to have type II error probability decreased in expense of type I error. Also, in a Bayesian approach all strategies with a p-value below 0.5 would have non-zero weights with higher weights for strategies with lower p-values (similar as in combining securities in Klein and Bawa (1976)). We take somewhat middle-ground with a threshold of 0.3 to filter the strategies that are good enough to be combined in the next step.

In the second step the best combination of the strategies that passed the first one is chosen. Here a conditional HM test is used. Essentially, each strategy is conditioned on every other strategy to check if there is value in combining them. The process is similar to the HM test except that instead of timing the market returns, we try to time the strategy returns with another strategy. For better clarity, we illustrate this with the following example. Let us assume the first strategy is the base strategy and the second one is the one to be tested for additional value. For every unique signal of the first strategy the median of market returns is calculated. For each unique signal a contingency HM table depicted by Table 14 is formed. For simplicity assume

Table 14: Example of a conditional HM test contingency table

	R > Median(R)	R < Median(R)
2nd signal cond. on 1st > 0	n11	n12
2nd signal cond. on 1st < 0	n21	n22

R is the market return, Median(R) is the median of the return.

that the table is for the signal of -1 of the first strategy. n11 is the number of times that the second signal is positive when the first signal is -1 and the market return is above the median return. If $\frac{n11+n22}{n12+n21} > 1$, the second strategy adds value to the first one when the first strategy's signal is -1. The overall conditional HM test contingency table is the sum of the tables for each unique signal of the first strategy. Thus, the N11 of the final table is the sum of all n11's, N12 is the sum of all n12's and so on. From such final table we can calculate:

- $N_{01} = N_{11} + N_{21}$
- $N_{02} = N_{12} + N_{22}$
- $N_{10} = N_{11} + N_{12}$
- $N_{20} = N_{21} + N_{22}$
- $N = N_{11} + N_{12} + N_{21} + N_{22}$

The HM statistic and the p-value are then calculated as usual:

$$HM = \frac{N_{11} - \frac{N_{10}N_{01}}{N}}{\sqrt{\frac{N_{10}N_{01}N_{20}N_{02}}{N^2(N-1)}}} \tag{29}$$

If the p-value is below 0.3, it is assumed that the second strategy adds value to the first one. If this is also true the other way around, both of the strategies are combined. If it holds only one-way, only the strategy that adds value to the other one is chosen. Since the underlying data for different strategies is not necessarily available for the same commodities or the same time period, the maximum overlapping signals sample is used for the comparison.

Finally, the choices of strategies across and within sectors are presented and their performance analyzed. A simple joint strategy that includes both across sectors and within sectors strategies is also described. There are cumulative performance graphs together with drawdowns and monthly returns presented. Furthermore, several performance measures for the chosen strategies are calculated with the focus on risk. These include IR, geometric IR, maximum drawdown, longest drawdown duration, maximum consecutive losing months, skewness, kurtosis and Ulcer Index. The Ulcer Index is a measure of risk based on the depth and duration of drawdowns proposed in Martin and McCann (1989). It is the square root of the average squared percentage drops from the previous peak and is calculated as follows:

$$UI = \sqrt{\frac{\Sigma_{i=1}^{N} {D_i}^2}{N}} \tag{30}$$

where

$$D_i = \frac{P_i - max(P_1, ..., P_i)}{max(P_1, ..., P_i)} \cdot 100 \tag{31}$$

The Ulcer Index can vary between 0 and 100 with higher number representing more risk in the performance. By construction, it is lower for strategies that avoid declines in portfolio value.

8.2 Results

The strategy results are presented and overviewed in this section following the methodology outlined in the previous section. First the best strategies that make bets on sectors are discussed and then we move on to the strategies within sectors. The section ends with an overview of the performance of the strategies that have been chosen.

In Table 15 the across sectors strategies' performance statistics are presented. Note that all results are based on active strategy returns on the benchmark (DJ-UBS Commodity Index). It can be observed that all strategies except Economic cycle sensitivity (number IX) have positive information ratios. However, if we look at the HM statistic and p-values, only 7 strategies out of 12 exhibit HM statistic's p-value smaller than the established threshold of 0.3. Momentum, roll yield, risk sensitivity and speculators' hedging pressure based strategies (III, IV, V, VI, VIII, X, XI) perform the best according to the HM test. Also, the expected probability of information coefficient being greater than 0.01 is above 0.5 for these 7 strategies.

Table 15: Strategies' results (across sectors)

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
IR	0.171	0.118	0.305	0.713	0.085	0.543	0.226	0.337	-0.079	0.353	0.279	0.082
GIR	0.042	0.028	0.079	0.192	0.018	0.146	0.057	0.090	-0.026	0.092	0.072	0.019
Skewness	0.011	-0.113	-0.386	-0.207	-0.239	0.155	0.358	0.355	-0.135	-0.052	-0.141	-0.275
Kurtosis	3.775	4.250	4.749	4.473	3.453	4.170	3.328	5.303	3.778	3.950	3.968	6.244
HM stat	0.193	0.453	0.586	3.560	0.578	2.668	0.128	0.953	0.137	0.857	1.029	-0.443
HM p-value	0.423	0.325	0.279	0.000	0.282	0.004	0.449	0.170	0.445	0.196	0.152	0.671
E[IC]	0.006	0.015	0.019	0.117	0.019	0.098	0.004	0.041	0.005	0.028	0.033	-0.020
E[P(IC>0.01)]	0.392	0.495	0.549	0.999	0.544	0.988	0.367	0.728	0.381	0.656	0.713	0.217
E[P(IC<-0.01)]	0.254	0.179	0.147	0.000	0.147	0.001	0.275	0.096	0.282	0.094	0.067	0.541

The columns represent different strategies as described in Section 8.1. I: 1-month momentum, II: 3-month momentum, III: 6-month momentum, IV: 12-month momentum, V: Roll yield, VI: Roll yield 5-year deviation, VII: Hedgers' hedging pressure, VIII: Risk sensitivity, IX: Economic cycle sensitivity, X: 12-month momentum with open interest growth confirmation, XI: Speculators' hedging pressure, XII: Overspeculation reversal. GIR is the geometric Information Ratio. Both IR and GIR reported are annualized. All results are based on active strategy returns on the benchmark (DJ-UBS Commodity Index).

To check which signals overlap and what is the best combination for the most predictive ability, we perform the conditional HM test described in Section 8.1. Note that for this test only the maximum overlapping sample of the chosen 7 signals is used in order to have a fair comparison. The results are presented in Table 16. The signals in columns are conditioned on the signals in rows. For example, in the first row, fourth column the p-value of 0.031 shows that strategy VI adds value to strategy III but III does not add value to VI (p-value of 0.692 in the fourth row, first column). In this case only strategy VI is considered. Again, threshold of 0.3 for the p-value is used. After inspecting Table 16, the conclusion is that strategy VI is superior with none of the other signals adding new information. Therefore, the final choice for the strategy across sectors is the one based on the 5-year monthly deviation of roll yields (number VI).

Table 16: Strategies' conditional HM test (across sectors)

	III	IV	V	VI	VIII	X	XI
III	0.500	0.346	0.276	0.031	0.152	0.186	0.446
IV	0.916	0.500	0.564	0.049	0.311	0.586	0.265
V	0.813	0.521	0.500	0.015	0.280	0.550	0.266
VI	0.692	0.547	0.541	0.500	0.554	0.702	0.327
VIII	0.733	0.226	0.370	0.041	0.500	0.211	0.312
X	0.627	0.233	0.353	0.012	0.163	0.500	0.281
XI	0.657	0.300	0.437	0.006	0.070	0.358	0.500

The rows and columns represent different strategies as described in Section 8.1. III: 6-month momentum, IV: 12-month momentum, V: Roll yield, VI: Roll yield 5-year deviation, VIII: Risk sensitivity, X: 12-month momentum with open interest growth confirmation, XI: Speculators' hedging pressure. The cells contain the conditional HM test p-value for the column strategy signal conditioned on the row strategy signal. For example, in the first row, fourth column the p-value of 0.031 shows that strategy VI adds value to strategy III but III does not add value to VI (p-value of 0.692 in the fourth row, first column).

Moving on to strategies within sectors, the results are presented in Table 17. As in previous table, all results are based on active strategy returns on the benchmark (DJ-UBS Commodity Index).

Table 17: Strategies' results (within sectors)

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
IR	-0.072	0.320	0.182	0.401	0.375	0.462	-0.009	0.041	-0.035	0.277	0.231	0.048
GIR	-0.024	0.083	0.045	0.105	0.098	0.124	-0.007	0.008	-0.014	0.072	0.058	0.010
Skewness	-0.022	-0.180	-0.174	-0.129	0.053	0.113	-0.182	0.015	0.142	-0.138	0.041	0.246
Kurtosis	3.005	4.275	3.593	4.012	3.067	3.613	4.228	5.659	3.607	5.143	3.867	7.317
HM stat	-0.213	1.190	1.566	1.897	0.463	1.128	-0.320	0.194	-0.032	2.558	0.599	-1.341
HM p-value	0.584	0.117	0.059	0.029	0.322	0.130	0.625	0.423	0.513	0.005	0.274	0.910
E[IC]	-0.003	0.018	0.024	0.029	0.010	0.028	-0.006	0.006	-0.001	0.065	0.012	-0.081
E[P(IC>0.01)]	0.111	0.583	0.726	0.829	0.400	0.698	0.140	0.393	0.165	0.976	0.439	0.056
E[P(IC<-0.01)]	0.215	0.013	0.005	0.002	0.117	0.040	0.331	0.254	0.182	0.001	0.086	0.863

The columns represent different strategies as described in Section 8.1. I: 1-month momentum, II: 3-month momentum, III: 6-month momentum, IV: 12-month momentum, V: Roll yield, VI: Roll yield 5-year deviation, VII: Hedgers' hedging pressure, VIII: Risk sensitivity, IX: Economic cycle sensitivity, X: 12-month momentum with open interest growth confirmation, XI: Speculators' hedging pressure, XII: Overspeculation reversal. GIR is the geometric Information Ratio. Both IR and GIR reported are annualized. All results are based on active strategy returns on the benchmark (DJ-UBS Commodity Index).

6 strategies out of 12 have HM p-values below 0.3 and thus are picked for further examination. These are 3-month momentum, 6-month momentum, 12-month momentum, roll yield 5-year deviation, 12-month momentum with open interest growth confirmation and speculators' hedging pressure (II, III, IV, VI, X and XI). All of them with the exception of XI feature the expected probability of information coefficient being greater than 0.01 above 0.5. To decide which combination of strategies to use, the conditional HM test is performed on the maximum overlapping sample of these 6 strategies. The p-values of the conditional HM test are reported in Table 18.

Table 18: Strategies' conditional HM test (within sectors)

	II	III	IV	VI	X	XI
II	0.500	0.499	0.489	0.404	0.026	0.816
III	0.583	0.500	0.549	0.678	0.029	0.902
IV	0.836	0.664	0.500	0.491	0.071	0.883
VI	0.825	0.705	0.430	0.500	0.024	0.921
X	0.940	0.877	0.789	0.676	0.500	0.901
XI	0.450	0.461	0.278	0.363	0.046	0.500

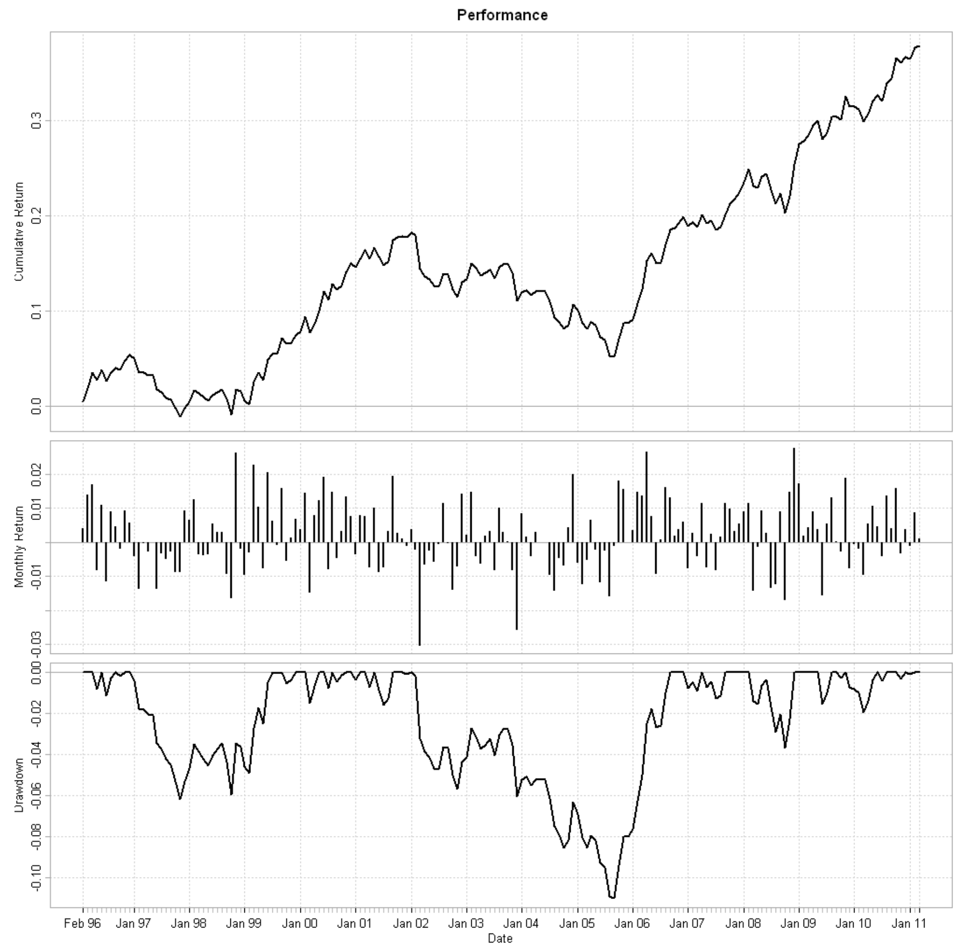
The rows and columns represent different strategies as described in Section 8.1. II: 3-month momentum, III: 6-month momentum, IV: 12-month momentum, VI: Roll yield 5-year deviation, X: 12-month momentum with open interest growth confirmation, XI: Speculators' hedging pressure. The cells contain the conditional HM test p-value for the column strategy signal conditioned on the row strategy signal. For example, in the first row, fifth column the p-value of 0.026 shows that strategy X adds value to strategy II but II does not add value to X (p-value of 0.940 in the fifth row, first column).

In the same fashion as in the table for strategies across sectors, the signals in columns are conditioned on the signals in rows. For example, in the first row, fifth column the p-value of 0.026 shows that strategy X adds value to strategy II but not vice-versa (p-value of 0.940 in the fifth row, first column). In this case only strategy X would be used. From Table 18, we draw the conclusion that the final combination consists of only the 12-month momentum with open interest growth confirmation strategy (X) as all others do not add new information to the signals.

The fact that both across sectors and within sectors have only one strategy in the end with others not adding extra value is in line with most of the literature discussed in Section 2. It seems that the strategies are driven by the same or similar underlying theory.

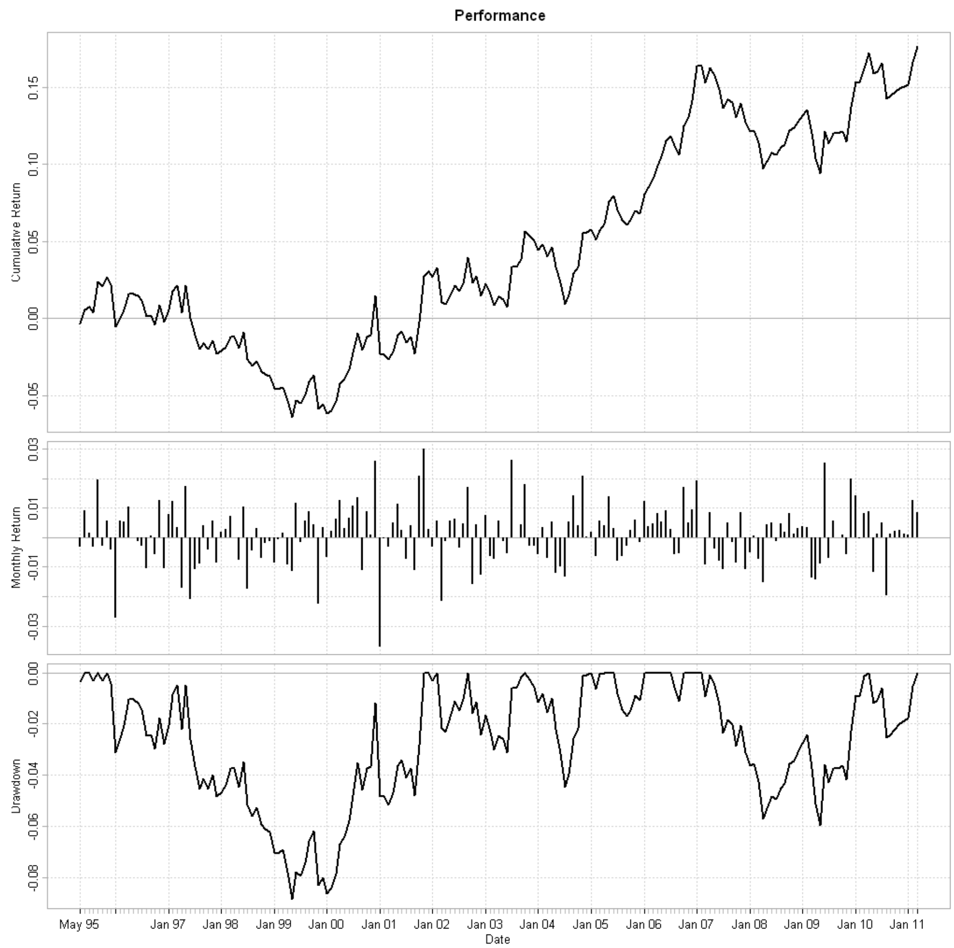
Next we present more detailed results of the strategies that have been picked. First, the cumulative performance graphs are presented for a quick overall picture and later more detailed numbers measuring performance are laid out. In Figure 3 the strategy VI across sectors cumulative returns are depicted together with monthly returns and drawdowns. Note that, as before, the strategy is scaled to have 1% historical volatility.

Figure 3: Strategy VI across sectors



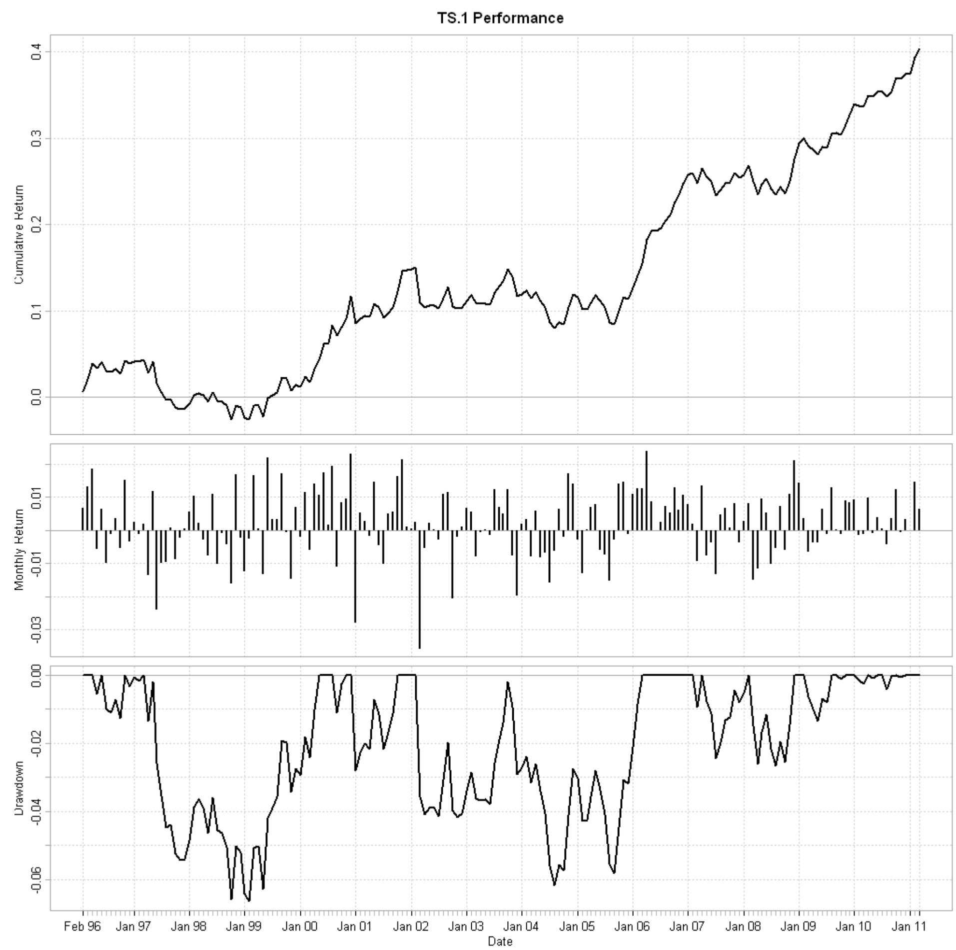
The performance of strategy VI across sectors varies over time with two major drawdowns that lasted 2.5 and 4.5 years. The last 5.5 years have been mostly successful with the strategy finishing 38% up for the whole period. A similar graph for the strategy X within sectors is presented in Figure 4.

Figure 4: Strategy X within sectors



The performance of strategy X within sectors seems more choppy, although the maximum drawdown is lower (9% versus 11%). The strategy has a cumulative return of 18% in the end of the period. Observing that the across sectors strategy and within sectors strategy have their biggest drawdowns in different periods, it could be a good idea to combine them. Indeed, the correlation is only 0.07. This is not surprising because one strategy takes bets on sector level while the other one stays neutral on sectors and bets on individual commodities within them. We take a look at the performance of these two strategies combined with simple equal weights. Again, the returns are rescaled for historical volatility of 1%. The graphs are depicted in Figure 5.

Figure 5: Combined strategy



As expected, the combined strategy has decreased drawdowns and therefore has smoother performance. The maximum drawdown lasting from 1997 to 2000 is slightly smaller than 7%. The strategy ends the period in March, 2011 with 40% of cumulative returns. More detailed numerical performance measures for both the across and within sectors strategies and the combination of them are reported in Table 19. Note that all strategies are scaled to have 1% historical volatility.

Table 19: Strategy performance

	Across sectors VI	Within sectors X	Combined
IR	0.5426	0.2769	0.5576
GIR	0.1462	0.0720	0.1497
Ulcer Index	3.6731	3.5277	2.7954
Max DD	0.1101	0.0887	0.0664
Longest DD (months)	55	71	48
Max c. losing months	11	5	4

Several performance measures for the two strategies and the combination of them. The first two rows are the Information Ratio and the geometric information Ratio. Ulcer Index is a risk measure based on the drawdowns (higher means more risk). DD in the fourth and fifth rows stands for drawdown. The last row is the maximum consecutive months with negative return.

The statistics confirm our observation in the graphs that the combined strategy is an improvement in risk over the separate strategies across and within sectors. While the IR and GIR are only marginally better than for the across sectors VI strategy, the four risk measures (Ulcer Index, maximum drawdown, duration of longest drawdown and maximum consecutive losing months) have improved. Our results are difficult to compare with the ones obtained by others because of several reasons. First, our strategies have slight modifications over the most simple ones. Second, the sample period differs. Finally, most of the strategies researched in the literature invest in futures contracts, not in indices. However, in a general sense they are in line with the positive results for roll yield and momentum based strategies in the previous literature.

To get a feeling for the positions taken by the strategies, we can take a look at Figures 6 and 7 for the across sectors and within sectors strategies respectively. Note that for the within sectors strategy only those commodities with available data for the strategy signal are reported. The green color depicts the periods where the signal was long, red stands for short positions and white is for neutral. In Figure 6 the data for Industrial Metals sector was not available until mid 2002, therefore a neutral position was held for that period. It is clear that no sector was dominantly long or short throughout the sample. Also, it seems that the positions change more frequently in the within sectors strategy than the across sectors one. From Figure 6 we see that the strategy was mostly long in precious metals since the end of 2007 and benefited from the growing silver and gold prices. Similarly the strategy was mostly short the energy sector since the end of 2008 (post sub-prime mortgage crisis and Lehman Brothers collapse) and profited from the slump in energy commodities.

Figure 6: Across sectors strategy positions

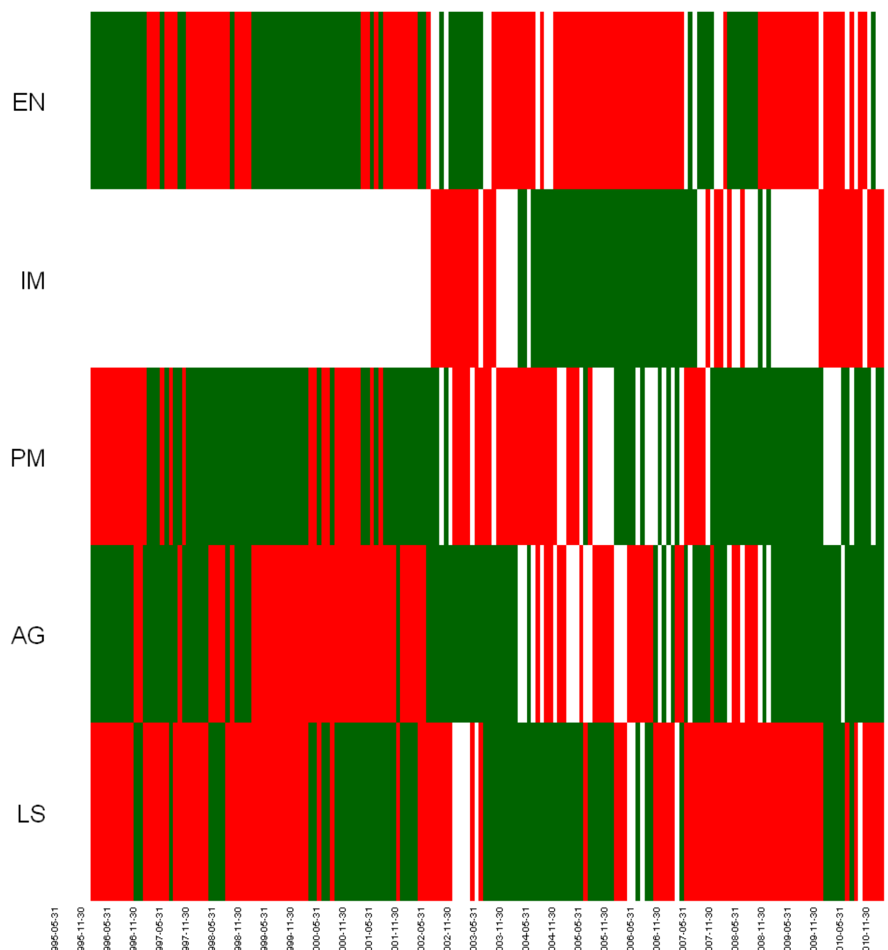


Figure 7: Within sectors strategy positions

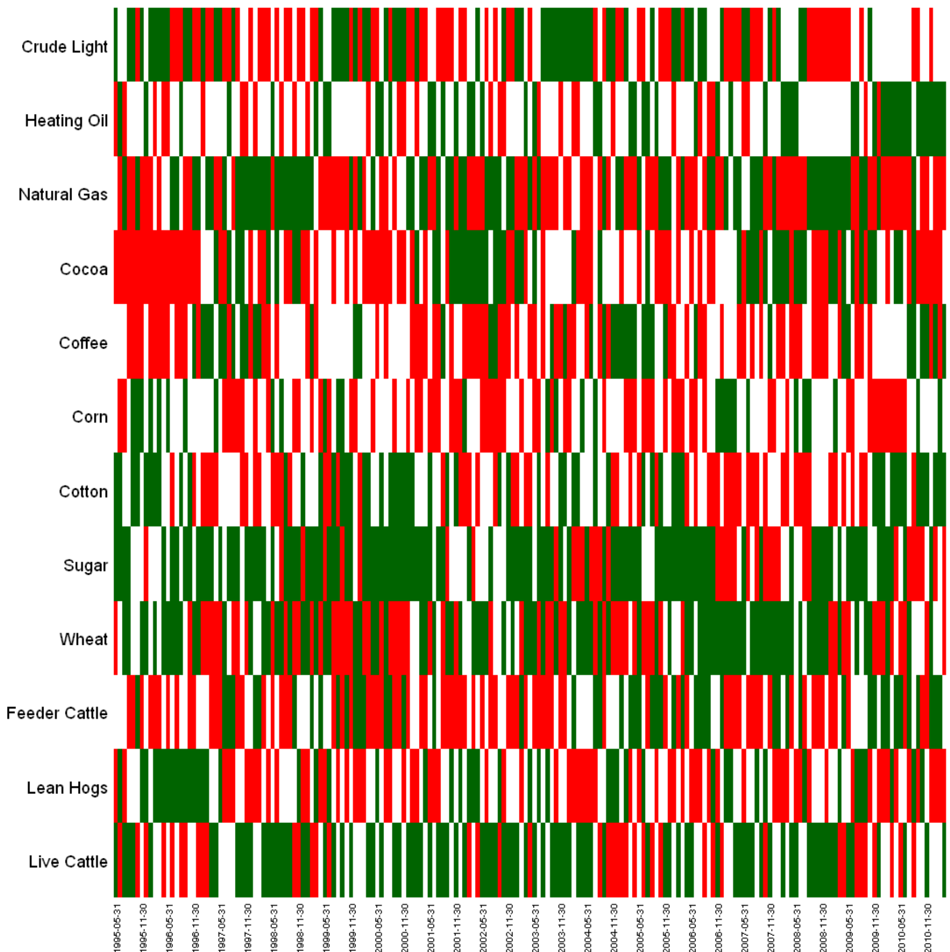


Table 20: Strategy positioning examples

	06/2008	12/2008	03/2011
Energy Sector	0.0849	-0.0849	0.0000
Industrial Metals Sector	0.0000	0.0000	-0.0849
Precious Metals Sector	0.0849	0.0849	0.0849
Agriculture Sector	-0.0849	0.0849	0.0849
Livestock Sector	-0.0849	-0.0849	-0.0849
Crude Light	0.0452	-0.0452	0.0000
Heating Oil	0.0000	0.0000	0.0452
Natural Gas	-0.0452	0.0452	-0.0452
Cocoa	0.0452	-0.0452	0.0000
Coffee	-0.0452	-0.0452	0.0452
Corn	0.0000	0.0000	0.0000
Cotton	-0.0452	0.0000	0.0452
Sugar	0.0000	0.0452	-0.0452
Wheat	0.0452	0.0452	-0.0452
Feeder Cattle	0.0452	0.0000	0.0000
Lean Hogs	-0.0452	-0.0452	-0.0452
Live Cattle	0.0000	0.0452	0.0452

The first five rows show the positions for the across sectors strategy while the others depict the positions for the within sectors strategy. The different columns represent the positions at different dates.

To be able to see the positions in more detail, we have taken three dates and presented the positions at those times in Table 20. June 2008 was when oil price was close to the peak, December 2008 the oil price had almost reached the bottom and March 2011 is the last observation in our research. In June 2008 the across sectors strategy was long the energy sector and the within sectors strategy went long crude light against a short in natural gas. As can be seen from Figure 6, there was a long position in the energy sector several months before June as well, meaning that the strategy profited from the growth in energy commodity prices. From Table 20 we see that by December 2008 the strategies' positions were reversed, with a short in the energy sector, long natural gas and short crude light. While the across sectors strategy lost because of the long positioning during part of the oil price crash of July 2008 - February 2009, it switched the signal successfully before the bottom and managed to profit from the last months of the slide. The across sectors strategy in March 2011 had a neutral position in energy, short in industrial metals and livestock and long in precious metals and agriculture sectors. The within sectors strategy was favoring heating oil, coffee, cotton and live cattle against natural gas, sugar, wheat and lean hogs.

9 Structural change tests

In this section we describe and perform structural change tests for strategies to check if there are any breaks in our sample period. Such a break could mean different performance in the sub-samples before and after the break. If a strategy fails one of the four tests, we date the estimated break and analyze what could have caused it. Depending on the cause, the strategy can be either modified or scrapped altogether. A good overview of different tests is provided in Zeileis (2005). We focus on the generalized fluctuation test framework. For the implementation we make use of the package *strucchange*[2] in the system for statistical computing *R*[3]. An excellent reference for the package is Zeileis et al. (2002).

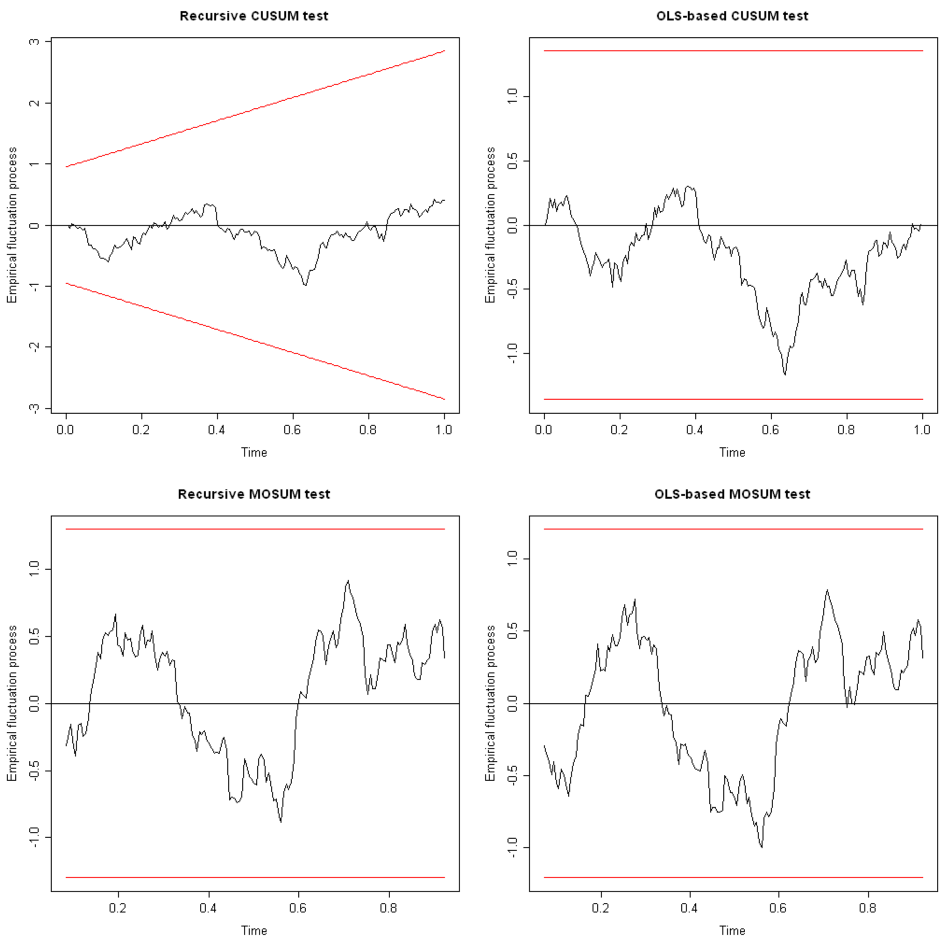
The general idea of structural change tests is based on a standard linear regression model. The structural change tests are performed on the strategy returns that are regressed on DJ-UBS Commodity Index total returns. The aim is to test if the null hypothesis of stable regression coefficients over time (no structural change) can be rejected. In a nutshell, generalized fluctuation tests derive an empirical fluctuation process from the fitted data that captures either fluctuations in residuals or in coefficient estimates. The limiting processes for these derived empirical processes are known, therefore, statistical significance boundaries can be computed. The null hypothesis is rejected if the derived empirical fluctuation process crosses one of the computed boundaries. We implement the tests based on the fluctuation in residuals. In this analysis, two estimates of the residuals are used: OLS and recursive. Also, two ways to estimate the fluctuation process are considered, one using cumulative sums of residuals (CUSUM) and the other one using moving sums of residuals (MOSUM) with the default bandwith size of 0.15 (meaning the window size is 15% of the whole sample). This gives us

[2] http://cran.r-project.org/web/packages/strucchange/index.html
[3] http://www.r-project.org/

in total four combinations of the empirical fluctuation process: cumulative sums of recursive residuals (CUSUM-REC), cumulative sums of OLS residuals (CUSUM-OLS), moving sums of recursive residuals (MOSUM-REC) and moving sums of OLS residuals (MOSUM-OLS). The limiting processes for CUSUM-REC and MOSUM-REC are based on the Brownian motion while limiting processes for CUSUM-OLS and MOSUM-OLS are based on the Brownian bridge. For more details and computations one can refer to Zeileis et al. (2002).

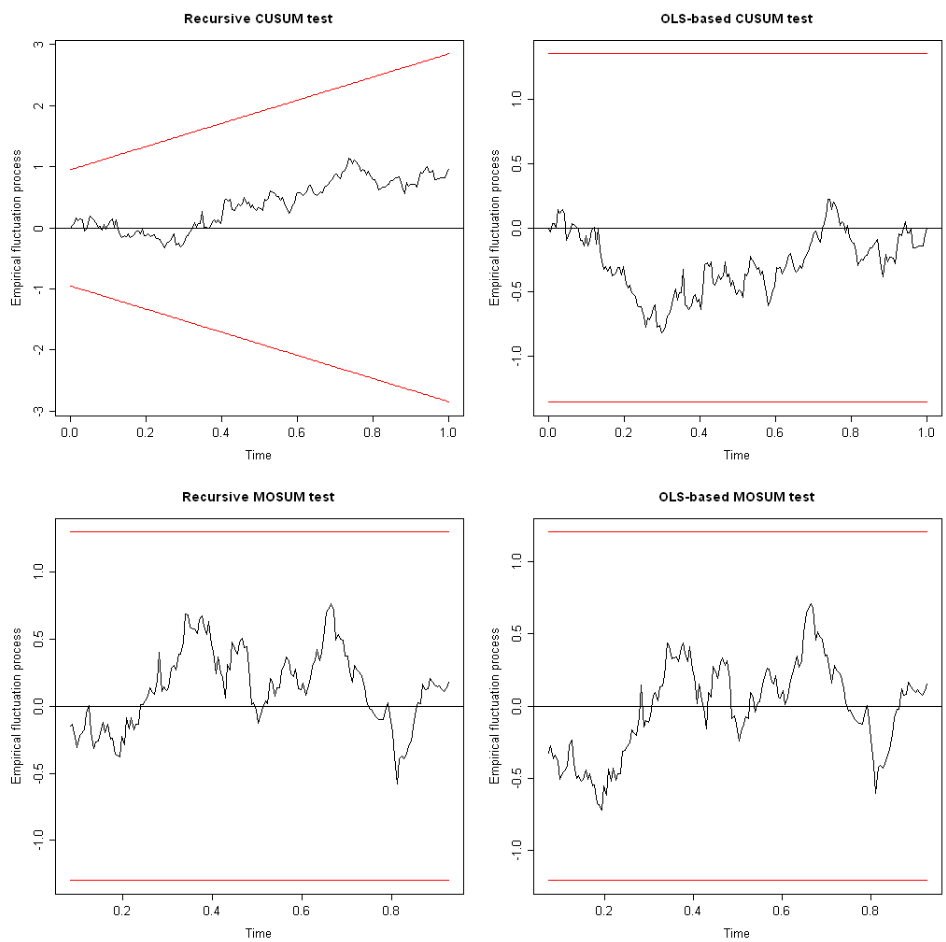
The plotted empirical fluctuation processes together with corresponding boundaries (at 5% significance level) for the strategy VI across sectors are presented in Figure 8.

Figure 8: Across sectors VI strategy structural change tests



It's clear that the tests do not reject the null hypothesis of no structural changes over the period. The same graphs are provided for the strategy X within sectors in Figure 9.

Figure 9: Within sectors X strategy structural change tests



Again, according to all four tests based on different empirical fluctuation processes, there is no reason to reject the null hypothesis at 5% significance level. The corresponding p-values of these tests are reported in Table 21. The lowest one is 0.13 for the cumulative sums of OLS residuals for strategy VI across sectors. Therefore, there are no indication that the behaviour of the strategies differs structurally over time and our results seem robust.

Table 21: p-values of structural change tests for strategies

	cusum-rec	cusum-ols	mosum-rec	mosum-ols
Across s. VI strat.	0.64	0.13	0.33	0.20
Within s. X strat.	0.72	0.51	0.44	0.42

10 Robustness checks

In this section we perform several robustness checks to see how sensitive our results are to the sample period for risk model estimation and to the arbitrary choices we have made for strategy performance.

10.1 Risk model

Due to computational time constraints, we do not follow through the whole stepwise regression procedure and only estimate the final risk model in two sub-samples to see how robust the results are. The sample is split in two sub-samples, 1991-2002 and 2003-2011. The preference was for evenly sized samples but open interest data for most of the commodities starts in 1995, therefore, the first sub-sample is a bit longer. The results of both sub-samples are presented in Table 22.

This table is compared to Table 9. The condition number is not reported as there was not enough data available for some of the multi-asset factors (see the list in Appendix F). We can see from Table 22 that the model would have been rejected in both sub-samples. In the first one the energy sector is not managable according to the characteristic portfolio test. The second one has agriculture sector, liquidity and materials sector equity factors that fail the test. However, the out-of-sample R^2 does not vary a lot. These results imply that one should be cautious when using the chosen risk model but we still think it is an improvement over using the market factor only. Note that the sub-samples are rather short due to the overall length of our dataset.

Table 22: Sub-sample analysis for the risk model

Sub-sample	R^2	p1	p2	p3	p4	p5	p6	p7	p8	p9
1991-2002	0.5799	**0.0305**	0.5048	0.7634	0.1354	0.0549	0.2625	0.5831	0.5569	0.9172
2003-2011	0.5537	0.2580	0.5202	0.6889	**0.0075**	0.7452	**0.0298**	**0.0202**	0.1423	0.7394

The first column differentiates between the model estimated in two sub-samples. The second one contains the out-of-sample R^2. Columns 3-11 contain the p-values of the characteristic portfolio beta test for every factor. p1-p5 are the p-values for the sector factors (energy, industrial metals, precious metals, agriculture, livestock), p6-p9 are the p-values for the liquidity, materials equity, 3-month momentum and the non-commercial hedging pressure factors in this order. The bold numbers indicate the p-values below the 5% threshold.

10.2 Strategies

It is required that the parameters chosen without solid fundamental basis should not be decisive in strategy's performance. There are two types of checks done. One is to modify the arbitrary parameters in a strategy and compare the new modified one with the original. The other test looks separately at the absolute strategy returns above and below their median. It is desired to have predictive power in signals that are associated with large absolute strategy returns. In Table 23 robustness check results for the roll yield deviation from 5-year average based strategy (number VI) across sectors are presented. The methodology for the calculation of the measures used is described in Section 8.1.

Table 23: Robustness check for strategy VI across sectors

	Orig.	3y	4y	6y	7y	Small	Big
HM stat	2.6677	3.6219	3.1577	2.2994	2.9186	1.2563	2.0948
HM p-value	0.0038	0.0001	0.0008	0.0107	0.0018	0.1045	0.0181
E[IC]	0.0984	0.1257	0.1129	0.0877	0.1177	0.0652	0.1088
E[P(IC>0.01)]	0.9884	0.9993	0.9970	0.9721	0.9947	0.8336	0.9648
E[P(IC<-0.01)]	0.0010	0.0000	0.0002	0.0035	0.0005	0.0612	0.0085

The first two rows are Henriksson-Merton test statistic and p-value. The expectations in the other rows are calculated using Bayesian methods as described in Section 8.1. The first column is the original strategy VI, the next four columns have 3-, 4-, 6- or 7-year windows respectively for the average roll yield instead of the 5-year window. The two last columns provide statistics for the signals that correspond to below and above median absolute strategy returns.

All four modifications to the average roll yield estimation window do not change the HM statistic significantly as can be observed from the first two columns. In addition, the expected probability of the IC being more than 0.01 for all of them is above 0.5. While the HM statistic for the signals that correspond to below median absolute strategy returns decreases from the original, the p-value is still below 0.3. More importantly, the last column displays that signals associated with above median absolute strategy returns have good predictive power with the HM statistic p-value of 0.0181. Therefore, the strategy indeed has predictive power when it matters the most.

Table 24 in the same fashion depicts the results for the 12-month momentum with open interest growth confirmation strategy within sectors.

Table 24: Robustness check for strategy X within sectors

	Orig.	11m	13m	1/4 buckets	Small	Big
HM stat	2.5578	2.4043	2.0977	2.5549	2.0919	1.6676
HM p-value	0.0053	0.0081	0.0180	0.0053	0.0182	0.0477
E[IC]	0.0653	0.0614	0.0536	0.1295	0.0752	0.0602
E[P(IC>0.01)]	0.9763	0.9663	0.9358	0.9882	0.9535	0.8955
E[P(IC<-0.01)]	0.0008	0.0013	0.0035	0.0022	0.0059	0.0184

The first two rows are Henriksson-Merton test statistic and p-value. The expectations in the other rows are calculated using Bayesian methods as described in Section 8.1. The first column is the original strategy X, the next two columns have 11-month and 13-month momentum respectively instead of the 12-month momentum. The fourth column shows the strategy results with long/short bucket size of 1/4 instead of 1/3. The two last columns provide statistics for the signals that correspond to below and above median absolute strategy returns.

The two modifications based on different amount of months for the momentum signal have no significant change in the HM-statistic and its' p-value. The same story goes for different bucket size. Finally, it is observed from the last two columns that both signals corresponding to below and above median absolute strategy returns have predictive ability. Also, for all modifications the expected probabilities of the IC being more than 0.01 are above 0.5.

As an additional check, we take a look at the correlation of the modified strategy signals with the original one. If they are positively correlated, it means that the good performance is indeed due to capturing the same effect. In Tables 25 and 26 the modified signal correlations with the original one are presented. All correlations are highly positive, implying that the slight

Table 25: Signal correlation for strategy VI across sectors

	3y	4y	6y	7y
Original	0.77	0.89	0.92	0.84

Table 26: Signal correlation for strategy X within sectors

	11m	13m	1/4 buckets
Original	0.96	0.96	1

modifications in strategies' parameters did not change the essence of the strategy.

Finally, we take a look at how the strategies performed in two sub-samples by splitting the full sample in halves. In Tables 27 and 28 the results for the across sectors and the within sectors strategies are presented respectively.

Table 27: Sub-sample performance for strategy VI across sectors

	Full sample	1st half	2nd half
HM stat	2.6677	0.6406	3.0878
HM p-value	0.0038	0.2609	0.0010
E[IC]	0.0984	0.0405	0.1394
E[P(IC>0.01)]	0.9884	0.6568	0.9971
E[P(IC<-0.01)]	0.0010	0.1905	0.0003

The first two rows are Henriksson-Merton test statistic and p-value. The expectations in the other rows are calculated using Bayesian methods as described in Section 8.1. The first column is the full sample results, the other two columns contain the results for the 1st half and the 2nd half of the sample respectively.

It seems that both of the strategies did much better in the second half of the sample. However, the HM p-value is below 0.3 and the expected probability of the IC exceeding 0.01 is above 0.5 in the first sample for both strategies. As there were no structural breaks indicated by the structural change tests in Section 9, we have no reason to suspect that this is something more than natural variation in the performance of the strategies. From these results it seems that both strategies are robust and perform as expected.

Table 28: Sub-sample performance for strategy X within sectors

	Full sample	1st half	2nd half
HM stat	2.5578	0.5915	2.5702
HM p-value	0.0053	0.2771	0.0051
E[IC]	0.0653	0.0249	0.0823
E[P(IC>0.01)]	0.9763	0.5930	0.9826
E[P(IC<-0.01)]	0.0008	0.1712	0.0011

The first two rows are Henriksson-Merton test statistic and p-value. The expectations in the other rows are calculated using Bayesian methods as described in Section 8.1. The first column is the full sample results, the other two columns contain the results for the 1st half and the 2nd half of the sample respectively.

11 Conclusion and further research

This paper analyzes risk model and trading strategies for investing in commodity indices. The study makes use of 27 Dow Jones-UBS commodity total return indices. We establish a stepwise regression procedure with several criteria to come up with the choice for the risk model. The time-series model consists of sector, liquidity, materials sector equity, 3-month momentum and non-commercial hedging pressure factors. It has an average out-of-sample coefficient of determination of 0.57. We find that a prominent theory that shows up behind variously constructed factors is the risk transfer from the hedgers to the speculators. We also raise some concerns about the model. First, the analysis shows that it is not clear if the liquidity factor indeed captures liquidity. Second, we perform Gibbons-Ross-Shanken asset pricing test and find out that there are deviations in commodity returns that are not explained by the model at 5% significance level. However, we discuss that this could be due to restrictive assumptions. Also, the sub-sample analysis shows that the model violates some of the criteria in shorter periods. Nevertheless, we think that the model is an improvement over having only the market factor that has the average out-of-sample R^2 of 0.23. The theories examined in risk model research and the insight gained is a good starting point for the strategy research which is the second major part of this paper.

We propose a list of strategies and describe the economic intuition behind each of them. Some of them have been researched in previous literature, while others have been proposed or modified in this paper. Each strategy is considered in two types of set up, one going long/short in sectors, the other going long/short in commodities within sectors. Both of them have a neutral position on the commodity market as a whole. We perform the Henriksson-Merton non-parametric test to establish which strategies have predictive ability. While several strategies have significant HM statistics, the test of strategies conditioned on one another within a set up type shows that they do not add value when combined. The best strategy in the across sectors set up is the strategy based on roll yield deviation from its' 5-year average and has an IR of 0.54. The corresponding HM test statistic is significant with the p-value of 0.004. Similarly, the best strategy in the within sectors set up is the 12-month momentum with open interest growth as confirmation. It displays an IR of 0.28 and the HM statistic p-value is 0.005. The structural change tests performed show that there are no significant structural breaks in the strategy returns. In addition, we carry out robustness checks that indicate that the performance of the strategies is not influenced by the arbitrary parameter choices. Finally, we show that the combination of these two strategies could have lower drawdowns as they have low correlation.

Our results might be difficult to compare directly to the rest of the literature but in a general sense they are in line with the positive results for momentum and roll yield based strategies obtained by others. The fact that we use commodity indices makes these strategies accessible to managers that have restrictions on investing in futures contracts. The positive performance and the conclusions from the structural change and robustness checks lead us to recommend implementing the strategies live.

There are several possible extensions of this study. The risk model estimation can be possibly improved. The first idea is to estimate Seemingly Unrelated Regression model as that would incorporate the cross-correlation between residuals of different commodities. Alternatively, a Bayesian approach employing a Bayesian Model Averaging technique could produce better results. The concerns raised about the model testing and restrictive assumptions could be partly solved by using Generalized Method of Moments test which is more flexible. As we have considered several strategies in the research, one could test for data snooping with White's Reality Check (White, 2000) in a similar fashion as in Sullivan et al. (2003). Other ideas on further strategy research include varying bet sizing as now the bet sizes are fixed throughout time and sectors. The current strategies are market neutral long/short. Therefore, combining them with a timing strategy of the broad commodity market could lead to a better overall performance.

References

Akram, Q. (2009). Commodity prices, interest rates and the dollar. *Energy Economics*, 31(6):838–851.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.

Basu, D. and Miffre, J. (2011). Capturing the Risk Premium of Commodity Futures: The Role of Hedging Pressure.

Baur, D. and Lucey, B. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. *Financial Review*, 45(2):217–229.

Baur, D. and McDermott, T. (2010). Is gold a safe haven? International evidence. *Journal of Banking & Finance*, 34(8):1886–1898.

Bessembinder, H. (1992). Systematic risk, hedging pressure, and risk premiums in futures markets. *Review of Financial Studies*, 5(4):637.

Bessembinder H. and Chan, K. (1992). Time-varying risk premia and forecastable returns in futures markets. *Journal of Financial Economics*, 32:169–193.

Borovkova, S. and Geman, H. (2006). Seasonal and stochastic effects in commodity forward curves. *Review of Derivatives Research*, 9(2):167–186.

Box, G. and Tiao, G. (1973). *Bayesian inference in statistical analysis*. Wiley Online Library.

Chang, E. (1985). Returns to speculators and the theory of normal backwardation. *Journal of Finance*, pages 193–208.

Chen, Y., Rogoff, K., and Rossi, B. (2010). Can Exchange Rates Forecast Commodity Prices? *The Quarterly Journal of Economics*, 125(3):1145.

Clark, T. and West, K. (2006). Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics*, 135(1-2):155–186.

Cleveland, R., Cleveland, W., McRae, J., and Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73.

Cochrane, J. (2000). *Asset pricing*. Princeton University Press.

Cootner, P. (1960). Returns to speculators: Telser versus Keynes. *The Journal of Political Economy*, 68(4):396–404.

De Grauwe, P. and Grimaldi, M. (2006). Exchange rate puzzles: a tale of switching attractors. *European Economic Review*, 50(1):1–33.

De Roon, F., Nijman, T., and Veld, C. (2000). Hedging pressure effects in futures markets. *The Journal of Finance*, 55(3):1437–1456.

Dow Jones Indexes (2010). *The Dow Jones-UBS Commodity Index Handbook*.

Durr, R. and Voegeli, M. (2009). Structural Properties of Commodity Futures Term Structures and Their Implications for Basic Trading Strategies.

Etula, E. (2009). Risk Appetite and Commodity Returns.

Fama, E. and French, K. (1987). Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. *The Journal of Business*, 60(1):55–73.

Fama, E. and French, K. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465.

Fama, E. and French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fisher, R. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, 1:3–32.

Froot, K. and Teo, M. (2008). Style investing and institutional investors. *Journal of Financial and Quantitative Analysis*, 43(04):883–906.

Fuertes, A., Miffre, J., and Rallis, G. (2010a). Strategic and Tactical Roles of Enhanced-Commodity Indices.

Fuertes, A., Miffre, J., and Rallis, G. (2010b). Tactical allocation in commodity futures markets: Combining momentum and term structure signals. *Journal of Banking & Finance*, 34(10):2530–2548.

Gibbons, M., Ross, S., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, 57(5):1121–1152.

Gorton, G., Hayashi, F., and Rouwenhorst, K. (2007). The fundamentals of commodity futures returns.

Grinold, R. and Kahn, R. (2000). *Active portfolio management: A quantitative approach for providing superior returns and controlling risk*. McGraw-Hill Companies.

Gyomai, G. and Guidetti, E. (2008). OECD system of composite leading indicators.

Hammersley, J. and Morton, K. (1956). A new Monte Carlo technique: antithetic variates. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 52, pages 449–475. Cambridge University Press.

Henriksson, R. and Merton, R. (1981). On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *The Journal of Business*, 54(4):513–533.

Hicks, J. (1939). *Value and capital: An Inquiry into Some Fundamental Principles of Economic Theory*. Claredon Press, Oxford.

Hong, H. and Yogo, M. (2009). Digging into commodities.

Hong, H. and Yogo, M. (2011). What Does Futures Market Interest Tell Us about the Macroeconomy and Asset Prices?

Kaldor, N. (1939). Speculation and Economic Stability. *Review of Economic Studies*, 7(1):1–27.

Keynes, J. (1930). *A treatise on money*. Macmillan, London.

Klein, R. and Bawa, V. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3(3):215–231.

Kruskal, W. and Wallis, W. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621.

MacKinlay, A. (1995). Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics*, 38(1):3–28.

Marquering, W. and Verbeek, M. (2004). The economic value of predicting stock index returns and volatility. *Journal of Financial and Quantitative Analysis*, 39(02):407–429.

Martin, P. and McCann, B. (1989). *The investor's guide to fidelity funds*. Wiley.

Miffre, J. and Rallis, G. (2007). Momentum strategies in commodity futures markets. *Journal of Banking & Finance*, 31(6):1863–1886.

Mori, H. (2004). Finite sample properties of estimators for the optimal portfolio weight. *Journal of the Japan Statistical Society*, 34(1):27–46.

Murphy, J. (1999). *Technical analysis of the financial markets*. New York Institute of Finance.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., and Giovannini, E. (2008). *Handbook on constructing composite indicators: methodology and user guide*. OECD publishing.

Pesaran, H. and Timmermann, A. (1994). A generalization of the non-parametric test of market timing. *Economics Letters*, 44.

Pirrong, C. (2005). Momentum in Futures Markets.

Richter, M. and Sørensen, C. (2003). Stochastic volatility and seasonality in commodity futures and options: The case of soybeans.

Roache, S. (2008). Commodities and the Market Price of Risk.

Shen, Q., Szakmary, A., and Sharma, S. (2007). An examination of momentum strategies in commodity futures markets. *Journal of Futures Markets*, 27(3):227–256.

Sørensen, C. (2002). Modeling seasonality in agricultural commodity futures. *Journal of Futures Markets*, 22(5):393–426.

Sullivan, R., Timmermann, A., and White, H. (2003). Scientific progress with data sharing. *International Journal of Forecasting*, 19:217–228.

Szymanowska, M., De Roon, F., Nijman, T., and Van Den Goorbergh, R. (2010). An Anatomy of Commodity Futures Risk Premiums.

Tang, K. and Xiong, W. (2010). Index investment and financialization of commodities.

Vrugt, E., Bauer, R., Molenaar, R., and Steenkamp, T. (2007). *Dynamic commodity timing strategies*. Risk Books, London.

White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5):1097–1126.

Working, H. (1949). The Theory of Price of Storage. *American Economic Review*, 39(6):1254–62.

Zeileis, A. (2005). A unified approach to structural change tests based on ML scores, F statistics, and OLS residuals. *Econometric Reviews*, 24(4):445.

Zeileis, A., Leisch, F., Hornik, K., and Kleiber, C. (2002). strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software*, 7(2):1–38.

A List of commodity indices

1. DJ-UBS Brent Crude Total Return Sub-Index
2. DJ-UBS Gasoil Total Return Sub-Index
3. DJ-UBS Crude Light Total Return Sub-Index
4. DJ-UBS Heating Oil Total Return Sub-Index
5. DJ-UBS Natural Gas Total Return Sub-Index
6. DJ-UBS Unleaded Gasoline Total Return Sub-Index
7. DJ-UBS Aluminium Total Return Sub-Index
8. DJ-UBS Copper Total Return Sub-Index
9. DJ-UBS Lead Total Return Sub-Index
10. DJ-UBS Nickel Total Return Sub-Index
11. DJ-UBS Tin Total Return Sub-Index
12. DJ-UBS Zinc Total Return Sub-Index
13. DJ-UBS Gold Total Return Sub-Index
14. DJ-UBS Platinum Total Return Sub-Index
15. DJ-UBS Silver Total Return Sub-Index
16. DJ-UBS Orange Juice Total Return Sub-Index
17. DJ-UBS Soybean Meal Total Return Sub-Index
18. DJ-UBS Cocoa Total Return Sub-Index
19. DJ-UBS Coffee Total Return Sub-Index
20. DJ-UBS Corn Total Return Sub-Index
21. DJ-UBS Cotton Total Return Sub-Index
22. DJ-UBS Soybean Oil Total Return Sub-Index
23. DJ-UBS Sugar Total Return Sub-Index
24. DJ-UBS Wheat Total Return Sub-Index
25. DJ-UBS Feeder Cattle Total Return Sub-Index
26. DJ-UBS Lean Hogs Total Return Sub-Index
27. DJ-UBS Live Cattle Total Return Sub-Index

B Gas Oil total return series comparison

Table 29: Datastram and DJ-UBS gas oil total return series comparison

Date	Datastream	DJ-UBS
...	...	...
10/1/2010	269.3018	540.8963
10/4/2010	272.0249	546.3726
10/5/2010	271.2737	544.8658
10/6/2010	274.748	551.8460
10/7/2010	270.147	542.6065
10/8/2010	270.4844	543.2865
10/11/2010	270.2034	542.7281
10/12/2010	268.1829	538.6715
10/13/2010	272.4604	547.2654
10/14/2010	271.8076	545.9558
10/15/2010	267.3303	536.9646
10/18/2010	**539.2183**	**539.2185**
10/19/2010	531.1641	531.1641
10/20/2010	533.4143	533.4144
10/21/2010	532.4795	532.4796
10/22/2010	529.6709	529.6711
10/25/2010	537.1714	537.1716
10/26/2010	536.9861	536.9861
10/27/2010	526.4956	526.4957
...	...	...

The second column contains the gas oil total return series from Thomson Reuters Datastream while the third column data is from the DJ-UBS website. The bold row indicates the switching pointfrom excess return to total return in Datastream series. The small differences after that are due to rounding errors.

C Commodity cumulative returns graphs

Figure 10: Energy sector cumulative returns

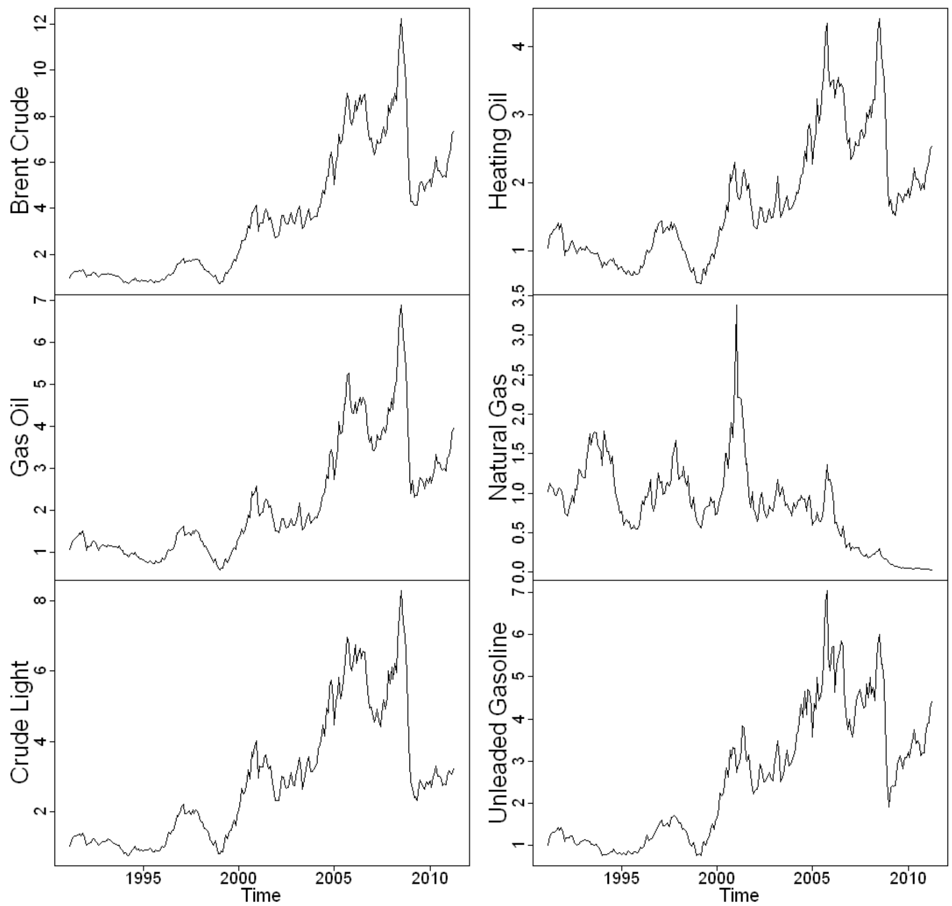


Figure 11: Industrial metals sector cumulative returns

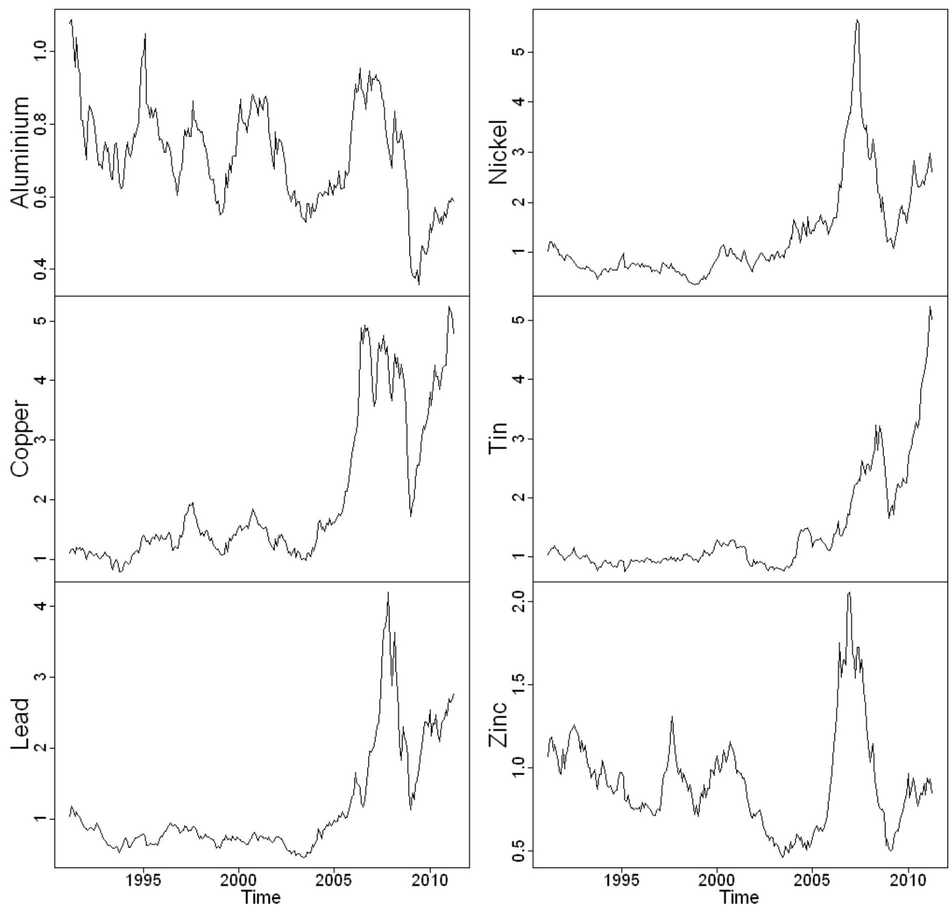


Figure 12: Precious metals sector cumulative returns

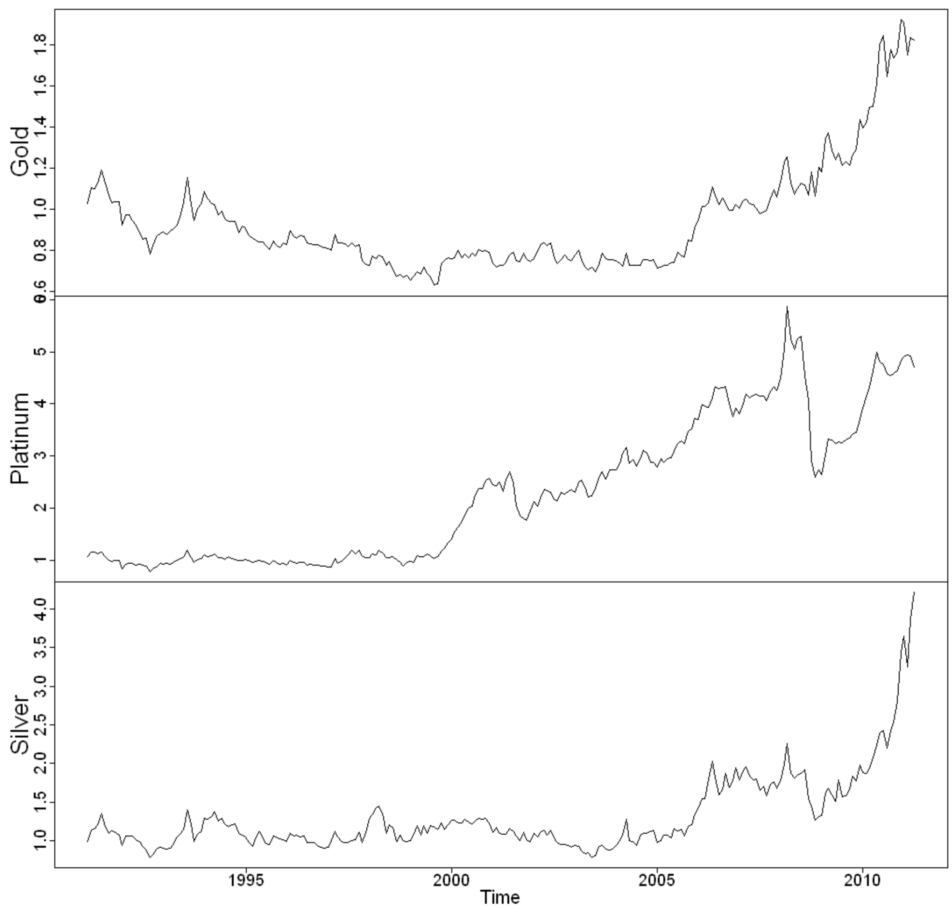


Figure 13: Agriculture sector cumulative returns

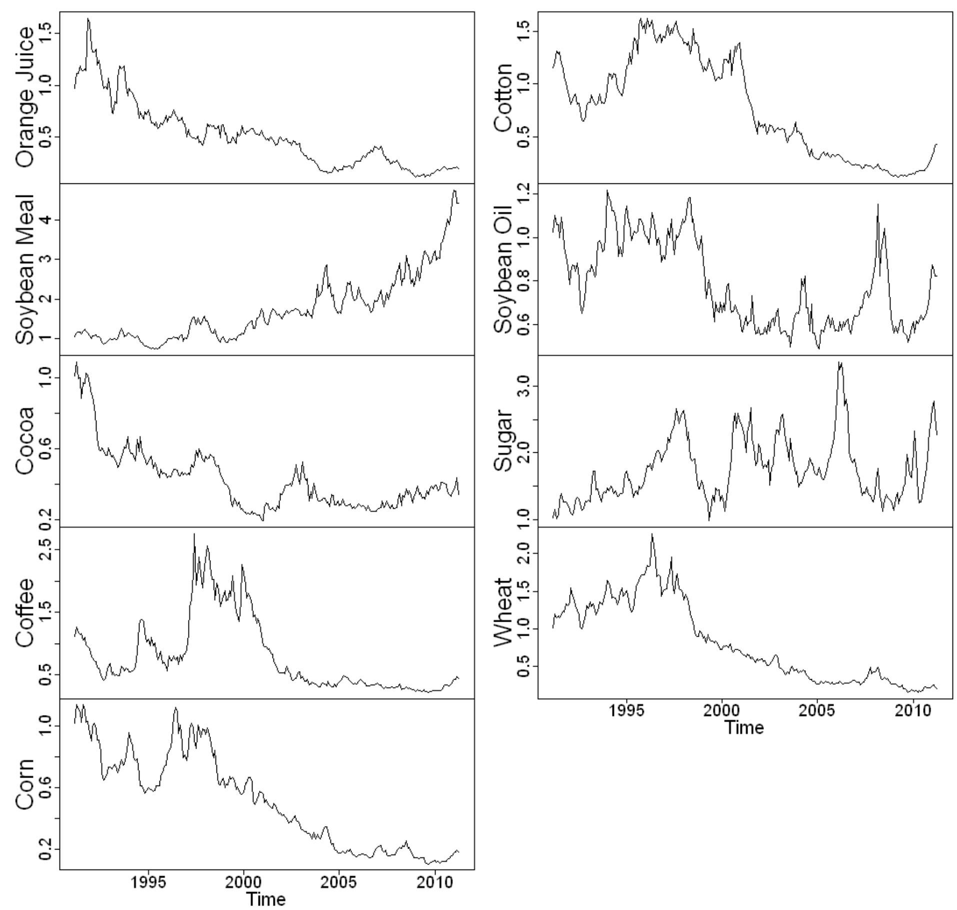
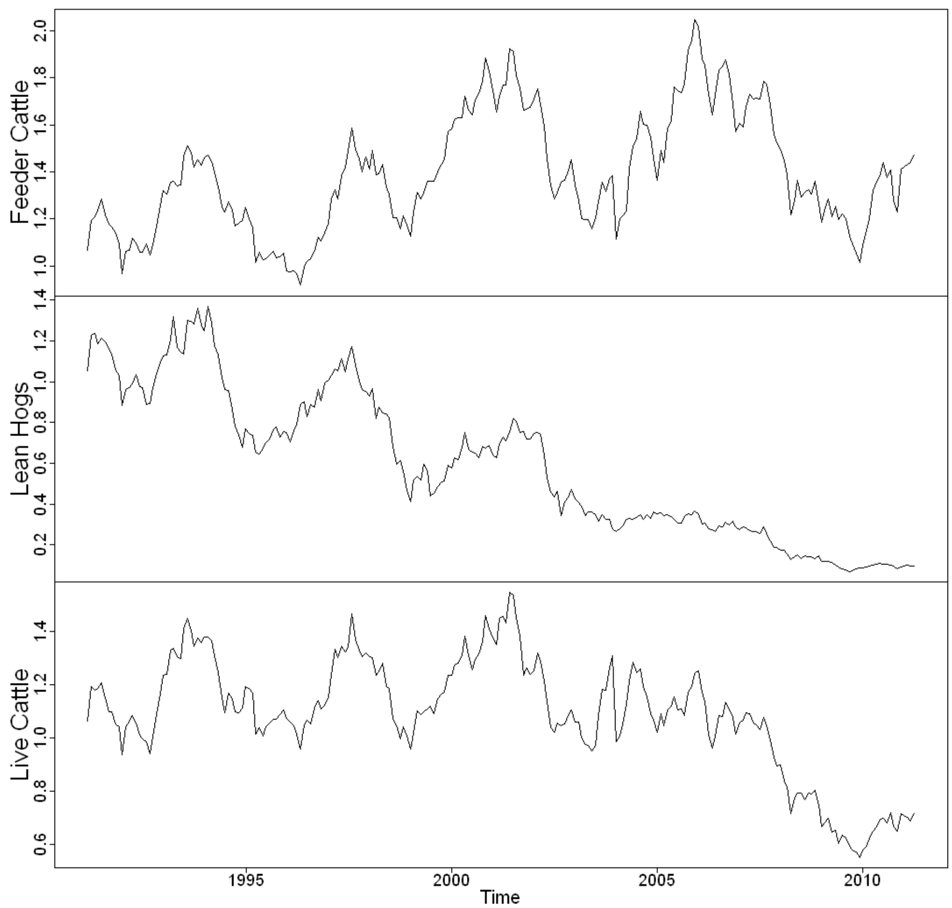


Figure 14: Livestock sector cumulative returns



D Commodities summary statistics

Table 30: Energy sector commodity returns

	Brent Crude	Gas Oil	Crude Light	Heating Oil	Natural Gas	Unleaded Gasoline
Min	-0.2723	-0.2970	-0.2662	-0.2607	-0.3454	-0.3274
Median	0.1266	0.1207	0.0834	0.1239	-0.0742	0.1622
Mean	0.1405	0.1140	0.1053	0.0937	-0.0534	0.1284
Max	0.3651	0.3459	0.3774	0.3642	0.4192	0.4069
StdDev	0.2877	0.3027	0.3066	0.3099	0.4841	0.3309
Skewness	-0.0399	0.0313	0.0240	0.2428	0.3366	0.0794
Kurtosis	4.6209	4.5014	4.2191	4.4444	3.2765	4.4371
ADF test	-9.1522	-9.3663	-9.3877	-9.8779	-11.5348	-10.8218

Min and Max are minimum and maximum monthly returns, while Median, Mean and StdDev rows all contain annualized values. ADF test row is the ADF test statistic. The ADF 5% significance critical value is -1.95, so a unit root is rejected for all series.

Table 31: Industrial metals sector commodity returns

	Aluminium	Copper	Lead	Nickel	Tin	Zinc
Min	-0.2181	-0.3000	-0.2764	-0.2906	-0.2054	-0.2702
Median	-0.0738	0.0865	0.0143	0.0147	0.0738	-0.0534
Mean	-0.0068	0.1122	0.0884	0.1063	0.1031	0.0226
Max	0.1501	0.2642	0.2534	0.3076	0.2382	0.2671
StdDev	0.1974	0.2620	0.2755	0.3456	0.2153	0.2491
Skewness	0.1353	-0.0099	-0.0311	0.2463	0.2145	0.1122
Kurtosis	3.8678	4.1539	4.1596	3.1193	4.5978	3.7630
ADF test	-9.7405	-9.0519	-9.9563	-9.8359	-9.1894	-9.1447

Min and Max are minimum and maximum monthly returns, while Median, Mean and StdDev rows all contain annualized values. ADF test row is the ADF test statistic. The ADF 5% significance critical value is -1.95, so a unit root is rejected for all series.

Table 32: Precious metals sector commodity returns

	Gold	Platinum	Silver
Min	-0.1155	-0.2830	-0.2171
Median	-0.0169	0.0841	0.0848
Mean	0.0420	0.0985	0.1088
Max	0.1568	0.1647	0.2236
StdDev	0.1572	0.2053	0.2743
Skewness	0.3081	-0.5587	0.0597
Kurtosis	3.7549	5.0807	3.0950
ADF test	-11.8162	-9.8881	-12.0773

Min and Max are minimum and maximum monthly returns, while Median, Mean and StdDev rows all contain annualized values. ADF test row is the ADF test statistic. The ADF 5% significance critical value is -1.95, so a unit root is rejected for all series.

Table 33: Agriculture sector commodity returns

	Orange Juice	Soybean Meal	Cocoa	Coffee	Corn	Cotton	Soybean Oil	Sugar	Wheat
Min	-0.2185	-0.2095	-0.2900	-0.2998	-0.1984	-0.1910	-0.2859	-0.2912	-0.1989
Median	-0.1319	0.0919	-0.1222	-0.1002	-0.0855	-0.0716	0.0265	0.0467	-0.1562
Mean	-0.0314	0.1074	-0.0101	0.0401	-0.0524	-0.0035	0.0238	0.0937	-0.0407
Max	0.4485	0.2540	0.3609	0.4988	0.2152	0.2519	0.2320	0.2946	0.2949
StdDev	0.3225	0.2604	0.2972	0.4114	0.2559	0.2791	0.2580	0.3271	0.2675
Skewness	0.7082	0.2192	0.6229	0.9675	0.0167	0.3242	-0.0251	0.1426	0.3661
Kurtosis	5.1326	3.5123	4.9670	5.0671	3.1512	3.2024	3.9266	3.3329	3.5984
ADF test	-11.1779	-10.7271	-12.2399	-9.9806	-9.2088	-9.8270	-11.1964	-10.6825	-11.3871

Min and Max are minimum and maximum monthly returns, while Median, Mean and StdDev rows all contain annualized values. ADF test row is the ADF test statistic. The ADF 5% significance critical value is -1.95, so a unit root is rejected for all series.

Table 34: Livestock sector commodity returns

	Feeder Cattle	Lean Hogs	Live Cattle
Min	-0.1963	-0.2616	-0.2477
Median	0.0433	-0.0968	-0.0186
Mean	0.0341	-0.0789	-0.0015
Max	0.1481	0.2591	0.1221
StdDev	0.1724	0.2684	0.1724
Skewness	-0.1250	-0.0522	-0.4275
Kurtosis	3.6421	3.3422	4.8867
ADF test	-10.0132	-10.4815	-10.5789

Min and Max are minimum and maximum monthly returns, while Median, Mean and StdDev rows all contain annualized values. ADF test row is the ADF test statistic. The ADF 5% significance critical value is -1.95, so a unit root is rejected for all series.

E Explanatory factors summary statistics

Table 35: Equity factors returns

	MSCI World	EM Eq.	Energy Eq.	Materials Eq.
Min	-0.1430	-0.1576	-0.1179	-0.1311
Median	0.0818	0.0258	0.0400	-0.0099
Mean	0.0469	0.0180	0.0301	-0.0080
Max	0.1175	0.1334	0.1138	0.1455
StdDev	0.1584	0.1567	0.1426	0.1196
Skewness	-0.4449	-0.2720	-0.0879	0.0676
Kurtosis	3.1336	3.5994	3.4323	4.9132
ADF test	-9.8472	-8.1551	-10.5578	-7.7771

The first column is the MSCI World index returns. EM Eq. is the emerging markets equity returns spread. The third and the fourth columns are energy and materials sectors equity returns spreads respectively. Min and Max are minimum and maximum monthly returns, while Median, Mean and StdDev rows all contain annualized values. ADF test row is the ADF test statistic. The ADF 5% significance critical value is -1.95, so a unit root is rejected for all series.

Table 36: Bond factors returns

	Long Rate	Term spread	Credit spread	Default spread	EM spread
Min	-0.0048	-0.0081	-0.0145	-0.0279	-0.0636
Median	0.0067	-0.0002	0.0012	0.0089	0.0171
Mean	0.0040	-0.0005	-0.0002	0.0046	0.0088
Max	0.0065	0.0055	0.0087	0.0239	0.0167
StdDev	0.0066	0.0057	0.0078	0.0223	0.0248
Skewness	-0.0894	-0.6627	-1.3797	-0.7959	-4.6268
Kurtosis	2.9355	6.3944	15.0864	7.3114	42.2772
ADF test	-10.7520	-11.0365	-9.9851	-9.6742	-9.0939

Columns 1-4 are self explanatory and described in detail in Section 3 while the last column is the emerging markets bond returns spread. Min and Max are minimum and maximum monthly returns, while Median, Mean and StdDev rows all contain annualized values. ADF test row is the ADF test statistic. The ADF 5% significance critical value is -1.95, so a unit root is rejected for all series. Emerging markets spread skewness and kurtosis values are so extreme mainly due to a few periods in 1998 (Asian crisis) and 2008 (Lehman Brothers collapse).

Table 37: Currency factors returns

	EURisk	CC-PC1	AUD	CAD	NZD	CLP	ZAR	NOK	ELMI
Min	-0.0590	-0.1156	-0.0649	-0.0745	-0.0810	-0.1031	-0.1039	-0.0890	-0.0601
Median	0.0121	-0.0228	-0.0169	-0.0330	-0.0015	0.0277	0.0265	-0.0071	0.0646
Mean	0.0019	0.0917	0.0023	-0.0126	0.0212	0.0450	0.1034	0.0068	0.0697
Max	0.0819	0.1841	0.0818	0.0932	0.1037	0.1168	0.1754	0.0958	0.0771
StdDev	0.0723	0.2023	0.0979	0.1054	0.1049	0.1286	0.1624	0.0712	0.0840
Skewness	0.0406	0.6415	0.4743	0.3562	0.5110	0.3809	0.8028	0.3533	-0.0586
Kurtosis	3.6848	3.3261	3.1226	3.1493	3.9994	3.5084	4.2623	7.3276	3.0877
ADF test	-11.5805	-7.8957	-8.2118	-8.7895	-8.4311	-8.6507	-8.2911	-8.6665	-9.4444

CC-PC1 is the commodity currency basket weighted according to the first principal component. Min and Max are minimum and maximum monthly returns, while Median, Mean and StdDev rows all contain annualized values. ADF test row is the ADF test statistic. The ADF 5% significance critical value is -1.95, so a unit root is rejected for all series.

F List of multi-asset factors

1. MSCI World index returns
2. MSCI Emerging Markets Europe index returns
3. MSCI Latin America index returns
4. MSCI Europe index returns
5. MSCI Japan index returns
6. MSCI Far East index returns
7. MSCI Pacific index returns
8. MSCI Consumer Discretionary index returns
9. MSCI Utilities index returns
10. MSCI Telecom index returns
11. MSCI Materials index returns
12. MSCI IT index returns
13. MSCI Industrials index returns
14. MSCI Health index returns
15. MSCI Energy index returns
16. MSCI Consumer Staples index returns
17. Equity size (small minus big) returns
18. Equity value (high minus low) returns
19. Long term interest rate
20. Slope of the yield curve
21. Barclays US High Yield Very Liquid index returns
22. JP Morgan Emerging Market Bond index returns
23. FTSE EPRA/NAREIT Global Real Estate index returns
24. EURisk: equally weighted EURUSD, EURGBP, EURJPY, EURAUD returns
25. JP Morgan Emerging Local Markets Plus index returns
26. Forex current account: long G10 currencies with current account surplus and short G10 currencies with current account deficit returns
27. Forex carry: long G10 currencies with high carry, short low carry G10 currencies returns
28. Purchasing Power Parity: long undervalued G10 currencies, short overvalued G10 currencies returns