Netspar

Network for Studies on Pensions, Aging and Retirement

NETSPAR ACADEMIC SERIES

Are commodity indices after their launch still rewarding for traditional portfolios?

Richard Jehnichen

MSc 08/2018-038

ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Master Specialisation Financial Economics

Master Thesis

Are commodity indices after their launch still rewarding for traditional portfolios?

Author: **Richard Jehnichen**
Student number: **458107**
Thesis supervisor: **Prof.dr. O. (Onno) Steenbeek**
First reader: **dr. R. (Ronald) Huisman**

Date final version: **30.08.2018**

Abstract

This thesis examines the diversification benefits of 1st, 2nd, and 3rd generation commodity indices for a traditional portfolio. A broad set of asset allocation strategies is employed in order to augment each benchmark portfolios that consists of 4 stock and 4 bond indices, with one of the seven commodity indices. This study examines the weekly return observations from February 2000 until May 2018, but due to backfilled data points, the main focus is on the actual performance starting in December 2010 until the end. The in-sample analyses, such as the spanning tests, and the other allocation strategies seem to exaggerate the benefits of commodities for portfolios as the out-of-sample fail to reproduce the results. First, 1st and 2nd generation indices fail to add value to a portfolio for the latest period. Second, heterogeneous performances of 3rd generation indices indicate that only a selection, but not all are superior to the previous generations or able to outperform the benchmark portfolio. Consequently, opposed to most of the previous studies, investments in 3rd generation commodity indices do not always yield beneficial performances.

Keywords: *Generation commodity indice, Asset allocation strategies, Spanning tests, Out-of-sample analysis, In-sample analysis,*

Table of Contents

List of tables

List of figures

CHAPTER 1

Introduction

The recent inclusion of commodity indices as portfolio diversifiers developed mixed conclusion. On the one hand, researcher present that commodity indices prove to be useful assets to diversify portfolios due to their low correlation with the traditional assets. On the other hand, some researcher already demonstrate that the increased interest led to the financialisation of this alternative asset class and weakened the potential benefits, especially after the financial crisis.

Recently, financial engineers constructed more sophisticated commodity indices with different investment strategies. The main differences is that the newest indices are now able to hold short investment positions in some commodity classes, whereas the older ones only invest in long positions. Therefore, researcher increasingly examine the different generation commodity indices in order to inspect the benefits of these in a traditional diversified portfolio (Belousova & Dorfleitner, 2012; Daskalaki, Skiadopoulos, & Topaloglou, 2017; Nijman & Swinkels, 2008; among others). However, since the newest and most advanced, the 3rd generation commodity indices launched only recently, they are especially still understudied. Daskalaki, Skiadopoulos, & Topaloglou (2017) and Kremer (2015) for instance are one of the first, who investigate the diversification performance of 3rd generation indices compared to the other generations. Their results indicate that the last generation perform superior to the other generations. These findings are confirmed and also rejected by Bessler & Wolff (2015) and Fethke & Prokopczuk (2018), who analyse the commodity indices on a benchmark portfolio consisting of only two assets as benchmark portfolio, based on monthly data. Next to that, Henriksen (2018) creates a more diversified portfolio, analyses weekly data points but only for the time period from 2000 until the end of 2015. Some of these indices however were only launch at the end of 2010. The main issue the previous researcher faced is that the data points from the time before the launch is backfilled by the index providers. The backfilled data however is only a hypothetical index value of the publisher, who might show biased performances to optimise their product. Contrary performances for 3rd generations are already presented by Fethke and Prokopczuk (2018), and consequently a more updated and extended analysis is needed in order to inspect the added value of the 3rd generation commodity indices to a portfolio for the period after their launch.

1.1 Research Questions and Hypotheses

Accordingly, the resulting two research questions can be stated as follows:

1. "*Do 3rd generation commodity indices add value to a traditional portfolio after their launch?*"

and

2. "*Do 3rd generation commodity indices outperform previous generations after their launch?*",

In the following, the four hypotheses can be formulated as:

$H0_a$: "3rd generation commodity indices do not add value to a traditional portfolio after their launch."

$H1_a$: "3rd generation commodity indices add value to a traditional portfolio after their launch.",

and also:

$H0_b$: "*3rd generation commodity indices do not outperform previous generations after their launch.*"

$H1_b$: "3rd generation commodity indices outperform previous generations after their launch.".

The answers to these questions and hypotheses are in particular interesting for real-world investors to provide them with more information about the potential benefits of especially 3rd generation commodity assets to a diversified portfolio.

Finally, this thesis contributes to the existing literature by focusing on the newest generation commodity indices and updating and extending previous studies. The framework of this research is from the point of view of an U.S. asset manager and empirically examines the advantages of 1st, 2nd and 3rd generation commodity indices to a diversified stock-bond portfolio, consisting of four stocks and four bonds. Moreover, the investigation is based on several sophisticated investment strategies and differentiates between in-sample as well as out-of-sample observations. The focus is on the last generation due to their recent launch and the more interesting investment strategies. The complete observation period is from February 2000 until May 2018 in order to examine a sufficient large data sample. Nevertheless, it concentrates on the period after the launch of the indices, starting in December 2010, since this is the only period with the actual performance and does not include backfilled data points that may potentially be misleading. The aim of this thesis is to shed further light on and to clarify the ongoing debate on the performance of generation commodity indices on a diversified equity-bond portfolio from an U.S. American perspective. The findings of this study are valuable for the investment decision of real world practitioners that try to explore more portfolio diversification opportunities.

The findings of this thesis depict that according to in-sample (IS) spanning tests, all indices are able to reduce the volatility of a diversified portfolio for every (sub-) period and some 3rd generation indices are even able to increase the portfolio return. Finally, since backfilled data of 3rd generation commodity indices seem to show overrated performances, the main focus is on the assessment of the after-launch period as it is more accurate appropriate than the previous or complete periods. Hence, the strategically IS analyses show that 3rd generations almost always lower the volatility but do only show inconclusive results for the after-launch period because of heterogeneous behaviours for different strategies. In general, 2nd generations outperform 1st generations and some 3rd generations outperform 2nd generation indices, which is in line with the literature. IS analyses however seem to overstate the returns and thus out-of-sample (OOS) examinations are more realistic.

For the OOS case, the assessment shows that there is a difference of the performance between generations and within generations. For the after-launch period, only some 3rd generation indices, namely the MSSO, the BBLS and the CDMNP are able to outperform a traditional portfolio, and present beneficial profiles for investors. All the other indices do not show favourable outcomes but are still able to lower the volatility. The results are in line with Fethke & Prokopczuk (2018), who presents heterogeneous 3rd generation indices. Nonetheless, they specify a different group of superior indices, which might be due to the investigation of longer periods that include backfilled observations and a less comprehensive benchmark portfolio.

The remainder of this paper is organised as follows. The second chapter discusses the existing literature and chapter three explains the methodology employed. Chapter four introduces the commodity indices and the database and the fifth chapter provides and discusses the descriptive statistics, the empirical results, and the limitations. In the last chapter, there is a short summary and finally a conclusion and discussion.

CHAPTER 2

Literature review

This chapter of the thesis describes the existing literature about portfolio efficiency. It describes the characteristics and fundamentals of commodity investments and the usefulness of the commodity inclusion in an investor's portfolio.

2.1 Strategic Allocation to Commodities

Greer (1978) and Bodie & Rosansky (1980) were one of the early researchers, who examined the risk and return of commodity future indexes for institutional investors and found that commodity derivatives yield the same mean return as common stocks and seemed to be good inflation hedges over the analysed period from 1950 to 1976. Next to that, they showed that individual commodities hardly exhibit significant correlation within each other.

Nevertheless, investments in commodities has been a relatively recent phenomenon. Since the last two decades, financial industry practitioners as well as academia increasingly emphasised strategic advantages of the inclusion of commodity related derivatives in portfolios next to the ordinary asset classes ((De Roon & Nijman, 2001; Jensen & Mercer, 2011; Lincke, 2016). Studies especially from the US showed that there exists a negative relation between US bond and stock markets, and commodity derivatives (Gorton, Rouwenhorst, & Bhardwaj, 2015). They appeared to be positively related with inflation rates and yield positive absolute returns, as well. Consequently, PF's started to invest part of their wealth in commodity indices in order to benefit from the risk diversification and inflation hedging attributes (Blitz & De Groot, 2014; Skiadopoulos, 2012). In the end, it seems that commodity derivative investments incorporate positive diversification effects and reduce return variability without forfeiting the returns of a traditional portfolio.

In 2006, Gorton & Rouwenhorst performed a comprehensive study covering a period of 45 years and examined the simple properties of commodity futures as an asset class. They discovered the fact that commodity futures historically also yield the same returns and Sharpe ratios compared to equities. Furthermore, they found that the risk premium is very similar to equities and that the return is negatively correlated with stock and bond markets but positively correlated with inflation and therefore provides positive diversification effects on a traditional portfolio with stocks and bonds. Hence, the hedging properties against unexpected inflation and the risk diversifying characteristics are one of the wide-spread reasons for investors to include commodities in their portfolio (Yau et al., 2007, p.526). In addition, "long-term investors with liabilities indexed to inflation, [like PF] may be able to improve their risk-return trade-off by including commodities in the portfolio" (Yau et al., 2007, p. 526).

Nonetheless, the implicit investment decision today is most vital for a fund's performance in the future. However, beneficial past performances do not necessarily imply favourable conditions in the future.
As can be seen already after the financial crisis in 2008, the previous growing interest and starting trend in alternative hedging vehicles for institutional investors diminished because of the simultaneously underperforming derivatives. Academics examined this relationship again and showed distorted commodity markets. Cheng & Xiong (2014) for instance examined the effects of the recently large inflows of investment capital to the commodities futures market and showed that the so-called financialisation significantly influenced the commodity derivatives market through distinctive risk sharing and information discovery mechanisms. The increased correlation between the alternative and traditional asset classes led to declining diversification options during that time and consequently to negative returns. Commodities even underperformed stocks in terms of the volatility and bonds in terms of the return (Blitz & De Groot, 2014). Next to that, they also confirmed that the correlation between the commodity market and the traditional markets increased in periods of volatile markets in combination with intense macroeconomic activity.

More recent literature nevertheless still argue that commodities add significant value to a portfolio (Lincke, 2016) and demonstrate that diversification and hedging benefits are still inherent in the portfolio inclusion of alternative asset classes (Daigler, Dupoyet, & You, 2017). In 2015, Gorton et al. (2015) updated their previous work for another ten years after the first publication in 2006 and also confirmed that most of the basic conclusions hold up out-of-sample, as well and commodities show similar returns and SRs than traditional asset classes.
Consequently, the motivation of this thesis is to further examine and clarify the current strategic asset allocation debate and the usefulness of more recent launched commodity indices as alternative asset class in a diversified portfolio.

Indexes are the most efficient way to invest in commodities (Greer, 2007). They allow an investor to acquire exposure to the broad commodities market without considering the underlying fundamentals (Miffre, 2012). Thus, several academic articles investigated the optimal investment allocation of institutional investors (Blitz & De Groot, 2014; Daigler et al., 2017; Nijman & Swinkels, 2003; Skiadopoulos, 2012; Yau et al., 2007). Belousova & Dorfleitner (2012) for instance studied the diversification effects of individual commodities from a European perspective and found that only energy commodities and precious metals add significantly more value to a traditional portfolio than soft commodities.[1] Consequently, one needs to distinct the types when investing in commodities since some of them exhibit more favourable conditions than others. Individual commodity classes are further

[1] Commodities exhibit similar but also wide-ranging investment characteristics. They are classified in two groups, namely the hard and soft commodities. The hard commodities include energy and metals, whereas the soft group contains agricultural, such as grains and seeds, and livestock products (Belousova & Dorfleitner, 2012).

corroborated by Skiadopoulos (2012), who showed that commodity indexes are not optimal approximations of an average commodity. Lincke (2016) also concludes that there are structural limitations of most of the major indexes and that investors should be active and knowledgeable when investing in alternative asset classes due to the large number of disparate and novel indexes available, nowadays.

The increased interest of institutional investors in commodity indices since the early 2000s engaged financial engineers to introduce numerous more commodity indices, which are already in its third generation, nowadays. Miffre (2012) was one of the first, who highlights the categorisation of commodity indices and examines the development of the three commodity indices and its performances. According to the categorisation, first generations indices are long-only positions that are passive and do not consider the market conditions. Second generation commodity indices are also long positions but adjust according to the current market and avoid contangoed but exploit backwardated conditions. A more detailed explanation is in Appendix 8.1. Finally, third generation indexes are different to the previous ones and are more flexible and follow active long-short positions based on (momentum) strategies.

Even though the new indices have been constructed in the recent years, empirical research about all types is still relatively scarce. Next to that, most of the available literature only examined in-sample periods, whereas real-time investors make decisions depending on out-of-sample periods (Daskalaki & Skiadopoulos, 2011).

Miffre (2012) is one of the first, who analyses all commodity index categories and looks at a time period of four years from 2008 until 2012. She concludes that the 2nd generation performance is superior to the first generation due to its characteristics to minimise contango conditions. The last generation however realises the highest performance for the lowest volatility and is the most beneficial category during extraordinary periods, in particular. The short analysis period however may be a limitation of such an analysis and thus Yan & Garcia (2017) extend the research and examined a period from 1991 until 2015. Furthermore, they compare the different generations of commodity indices and base their portfolio on the view of an U.S. investor. Their results also indicate that first and 2nd generation commodity indices do not have a significant effect on a portfolio's Sharpe ratio during the period. However, the evidence also indicates that the 3rd generation indices significantly add value to a portfolio. In their research, only the 3rd generation commodity index improves portfolio performance and either increase return and/or decrease risks. The evaluation of only one single index of each generation may however be a shortcoming as it might fail to represent the variety of these financial derivative (Fethke & Prokopczuk, 2018). Then, Kremer (2015) expanded the analysis and incorporated more indexes. The study also discovered diversification benefits of all generation indices for an U.S. investor out-of-sample portfolio from 1991 until 2013. He found in- as well as out-of-sample evidence for advantageous characteristics

of only third generation indices due to lower volatility and higher returns in a traditional portfolio but not strong beneficial results for the previous generation indices. Daskalaki, Skiadopoulos, & Topaloglou (2017) use a stochastic-dominance efficiency approach to identify the added value of commodities in a diversified portfolio. They summarise that indices, which imitate dynamic trading strategies show stronger diversification benefits during in- but also out-of-sample periods.

Henriksen (2018) extended the previous literature in using different improvement measures, employing four different long/short indices and creating a more diversified stock and bond portfolio. His conclusion indicates that third generations in general significantly improve the risk-adjusted performance measure besides divergent performances for separate sub-periods. During the back-testing periods for instance, the results show positive performances prior to their launching, but negative effects during their release period, which the authors could not explain and left open for further research.

In another recent article from Fethke & Prokopczuk (2018), all previous hypotheses about commodities in portfolios are confirmed and evidence for superior characteristics of long/short generation indices to diversify portfolios over the other two generations are shown. Nevertheless, they also emphasise that the applied third generation indices from 2000 until 2017 exhibit heterogeneous properties and surprisingly there are 3rd generation indices that adds value to a diversified portfolio, but also some that do not add value to a portfolio. In the end, this study builds upon this research and tries to extend and update their approach by using a more extensive portfolio, different return data and a more appropriate analysis period.

Summarising, even though there has been a lot of controversial literature available about the performance of commodity indices in general on a diversified portfolio, few studies look at the usefulness of the newest generation of indices. Even though the most recent articles argue that some third generation commodity indices show better performances than older generation indices this thesis tries to challenge their findings by approaching this study with a different set of analysis tools. Several previous studies for instance mainly use two stocks and bonds indexes to create a diversified portfolio (Daskalaki et al., 2017; Kremer, 2015; Yan & Garcia, 2017). This approach however, is for several reasons not appropriate to build a well-diversified portfolio and thus this thesis tries to extend it by using a more comprehensive benchmark portfolio.

2.1.1 Strategic timing decision to invest in commodities

As shown in the previous section, there are more and less beneficial periods for investors to invest in commodity indices. In general and according to the roll return component, a trading strategy yields positive returns and a higher Sharpe ratio than a long-only position if it is able to go long in commodity index during backwardated periods and short during contangoed ones (Erb & Harvey, 2016; Gorton &

Rouwenhorst, 2006). In another study by Miffre & Fernandez-Perez (2015), they confirm that long-short strategies of commodities depict superior risk-adjusted portfolio performances. Further, in particular during volatile markets, the long-short strategy correlates less with traditional assets and shows a lower volatility than long-only commodity portfolios. Nijman & Swinkels (2008) also find empirical evidence that active tactical trading decisions perform superior to fixed allocations for pension funds. Finally, in a paper by Miffre (2015), where she summarises the prevalent literature about long-short or long-only investment strategies in commodities, she also notes the supremacy of long-short positions in commodity futures market over long-only ones.

Another intriguing question to answer is the performance and behaviour of commodity indices during different business cycles, also compared to the traditional asset classes.
Gorton & Rouwenhorst (2006) and Bjornson & Carter (1997) analysed that commodity indices show different performances under different economic cycles and monetary policies. Furthermore, they proved that commodity futures perform better in periods of unexpected inflation and diversify the variation in traditional portfolio returns. Next to that, Yau et al. (2007) show that the correlation between commodities and the traditional asset classes seem to be time-period dependent. They found that during certain time periods, the correlation of commodities and other asset classes varies greatly but is in general neutral or even negative. This further emphasises the notion that commodities are business-cycle sensitive, which should imply an active attitude towards investments in this asset class.

CHAPTER 3

Methodology & hypothesis building

This chapter presents the applied methodology and comes up with more precise and updated hypothesis. Most of the methodology is in line with the recent asset management developments in academic literature that examine the usefulness of commodity indices as strategic portfolio allocation.

3.1 In-sample (IS) approach

3.1.1 Mean-variance portfolio

The most widely accepted theory in portfolio analysis, which is employed by most of the academic literature that analyses strategic portfolio allocations, is the Markowitz mean-variance (mv) optimisation strategy (Daigler et al., 2017; Nijman & Swinkels, 2003; Yan & Garcia, 2017; amongst others).
The fundamental idea behind this theory states that a portfolio is mean-variance efficient or optimised if there is no other portfolio available that yields higher returns with the same risks and combination of assets. Stated differently, diversification effects across more asset classes lead to a portfolio with lower risks or variances but the same returns (Markowitz, 1952).

3.1.2 Mean-variance spanning test

Nowadays, academia who study the effects of the inclusion of assets on a traditional portfolio however use an updated and modified approach of the so-called mean-variance and step-down spanning test based on Kan & Zhou (2012). Spanning tests, which were firstly introduced by Huberman & Kandel (1987), check if the efficient frontier of a benchmark portfolio, can be improved by including an additional test asset. So it tests if an additional investment can statistically and significantly improve the efficient frontier of an existing portfolio due to diversification benefits. The underlying assumption of these methods is that an investor obtains utility only from the mean and variance of returns (Fethke & Prokopczuk, 2018). Spanning is achieved if the efficient frontiers of K risky assets coincide with the efficient frontier of an augmented set of N plus K risky assets, where N symbolises the number of additional/test assets, K the number of benchmark assets (the portfolio) and T the number of time-series observations. In the end, spanning tests analyse the performance of the additional asset of in-sample (IS) portfolios without rebalancing (Bessler & Wolff, 2015).
In this thesis, the ideas of Kan & Zhou (2012) are employed to analyse the mean-variance spanning test. Since the selected test assets are regressed against the benchmark assets, equation (1) is the spanning regression:

$$R_{2t} = \alpha + \beta \cdot R_{1t} + \varepsilon_t \; ; \qquad t = 1,2,\ldots,T, \qquad (1)$$

$$R_{Commodity} = \alpha + \beta \cdot R_{Portfolio,t} + \varepsilon_t \; ; \qquad (2)$$

where R_t [R'_{1t}, R'_{2t}] depicts the excess return of N + K at time t, R_{2t} is a vector of the N test asset returns, R_{1t} a vector of the K benchmark asset returns, ε_t is the vector of the error terms, and T represents the number of time series observations, α is the interception value of the regression and β ($\beta = \beta_1,.., \beta_K$) is the slope of the regressions. In this study, N = 1 because only one commodity index is included in the portfolio regression per time.[2] R_t is defined as excess returns and not as total returns as Belousova & Dorfleitner (2012) emphasise, because excess returns are econometrically and for time-series analysis more appropriate when a risk-free asset is added to a portfolio (Xiong, Ibbotson, Idzorek, & Chen, 2010). Further underlying assumptions of spanning tests are that the expected value of the error term is zero ($E(\varepsilon) = 0_N$) and that the error terms are uncorrelated ($E(\varepsilon R') = 0_{N*K}$).

Consequently, the null hypothesis of spanning test can be stated as follows:

$$H_0 : \alpha = 0_N \, , \, \delta = 0_N; \qquad (3)$$

where $\delta = 1 - \beta 1_k$ with 1_k as a vector of ones.

As can be seen, equation (3) is a joint hypothesis and states that the null hypothesis of the test can be rejected if (1) the test asset (or portfolio of test assets) can statistically and significantly improve the augmented efficient frontier of the initial asset universe ($\delta \neq 1 - \beta = 0$), which is also called Global Minimum Variance Portfolio (GMVP), and also if (2) the inclusion of the additional asset enhances the tangency portfolio ($\alpha \neq 0$) (Belousova & Dorfleitner, 2012; De Roon & Nijman, 2001; Fethke & Prokopczuk, 2018; Kremer, 2015). A more detailed description and solution of the joint tests is given in section 3.1.4.

3.1.3 Additional spanning test under non-normally distributed returns

The standard spanning test however is not the most applicable method in this context and therefore, several more complementary forms of spanning test needs to be considered (Fethke & Prokopczuk, 2018).

The three standard extensions of spanning tests are the *Wald* (W), the *Lagrange Multiplier* (LM) and the *Likelihood Ratio* (LR) test, which are based under the assumption of normally distributed returns and are estimated by using the maximum likelihood estimator (MLE) (Belousova & Dorfleitner, 2012; Fethke & Prokopczuk, 2018; Kan & Zhou, 2012; Kremer, 2015). The null hypothesis of each additional test says that the inclusion of the additional test assets is spanned by the initial asset portfolio (see equation (3)). Accordingly, the null hypothesis can be rejected if it is possible to improve the efficient

[2] For more details of the general idea and regression with different values for N, see Kan & Zhou (2012).

frontier of the investment portfolio with the inclusion of a new test asset. Even though these tests can be ranked in finite samples as they exhibit conflicting findings (Berndt & Savin, 1977; Breusch, 1979), Belousova and Dorfleitner (2012) argue that all tests need to be performed to produce reliable results. More detailed descriptions of the formulas can be found in Appendix 8.2.

Next to that however, Erb & Harvey (2016) and Jensen & Mercer (2011) demonstrate that commodity future returns are not normally distributed and show conditional heteroskedasticity. Consequently, the previous three test statistics will not show adequate and valid test results under the basic formula, since they are not asymptotically chi-squared distributed (Belousova & Dorfleitner, 2012; Kan & Zhou, 2012; Kremer, 2015).

The assumptions that returns are normally distributed and the error terms in (2) are homoscedastic are the foremost shortcomings of the MLE (Kremer, 2015). Thus, the MLE-based approach in the three spanning test is replaced by the *Generalised Method of Moments* (GMM) methodology, developed by Hansen (1982). The GMM is beneficial because it adjusts for the non-normality assumption of asset returns (Belousova & Dorfleitner, 2012; Fethke & Prokopczuk, 2018; Kan & Zhou, 2012; Kremer, 2015). For this study, the sole generation of the Wald test results in combination with the GMM is sufficient enough, as the LM as well as the LR show the same results (Newey & West, 1987).

3.1.4 Step-down spanning tests

As described in section 3.1.2, the mv spanning tests depict a joint hypothesis of which the results are however more influenced by and biased towards the GMVP ($\delta = 0$) due to the statistically easier detection but to a lesser extent influenced by the tangency portfolio ($\alpha \neq 0$), even though it might be economically more important. This is because spanning test in general are more influenced by assets that lower the variance of the GMVP and less influenced by assets that improve the tangency portfolio (Kremer, 2015). Therefore, Kan & Zhou (2012) recommend a sequential two-step process to solve this issue. Moreover, it finds the source that rejects the null hypothesis. Two new F-test null hypotheses are designed by splitting the previous null hypothesis, namely:

$$H_{0,}^{F1} : \alpha = 0 \tag{4}$$

$$H_{0,}^{F2}{}_{|\alpha = 0} : \delta = 1 - \beta = 0 \tag{5}$$

The acceptance of equation (3) indicates that the two tangency portfolios are statistically similar. Equation (4) states that that the GMVP is statistically similar but is conditional and only holds if equation (3) can be accepted. The two F-tests are formulated as:

$$F_1 = (T - K - 1) \cdot \frac{\Sigma''}{\Sigma' - 1} \tag{6}$$

$$F_2 = (T - K) \cdot \frac{\Sigma`}{\Sigma"-1} \tag{7}$$

and where ∑" equals the unconstrained and ∑' the constrained estimator of MLE, when α=0. The ∑´ is the constrained estimator if α=0 and δ = 0 hold. For the null hypotheses both F-tests follow the F-distribution but the F_1-test follows with 1 and (T-K-1) degree of freedom, in comparison to the F_2-test that follows with 1 and (T-K) degree of freedom (Kremer, 2015).

Finally, the usage of the step-down approach is necessary to produce reliable and accurate information about the effects of commodities on a traditional portfolio and is superior to the three standard mv spanning tests (Kan & Zhou, 2012). First, it allows to discover the cause of a rejection of the null hypothesis either due to a raised tangency portfolio, or due to an enhanced GMVP. Second, the significance level can be selected more flexible depending on the economic significance of the two influenced elements.

3.1.5 The in- and out-of-sample settings and further adjustments

Since the in-sample (IS) analysis looks at an estimation window and evaluates the strategy according to that same window given perfect forecasts, for the out-of-sample (OOS) analysis, the strategy is assessed to the next window shifted one week forward and not given perfect forecasts. Especially real-time investors are more concerned about the OOS ex-ante benefits of test assets in the ex-post setting (Kremer, 2015; Skiadopoulos, 2012).

Previous articles exhaustively studied the diversification effects of commodities from an in-sample approach (Belousova & Dorfleitner, 2012). Only more recently researcher look at the OOS setting (Daigler et al., 2017; Daskalaki et al., 2017; Fethke & Prokopczuk, 2018; Henriksen, 2018; Yan & Garcia, 2017). Bessler & Wolff (2015) for instance, conclude that in-sample analyses show positively biased results for diversification effects of commodities. Hence, the out-of-sample approach shows more realistic results for real world investor since an asset manager decides about the asset allocation strategy of the portfolio from the information derived at that moment in time (= in-sample). The evaluation of the investment returns is out-of-sample however. The outcome is based on the optimised in-sample strategy that is only realized at the end of her investment horizon in her future (= out-of-sample).

3.2 Out-of-sample (OOS) approach

3.2.1 The OOS and rolling-sample estimation procedure

The OOS theory is based on DeMiguel, Garlappi, & Uppal's (2009) study and a "rolling-sample" approach is applied to construct the OOS asset allocation strategy. This method is appropriate to analyse the OOS results and is also employed in several more related asset allocation articles such as in Bessler & Wolff (2015), Fethke & Prokopczuk (2018) and Kremer (2015).

In particular, it considers 956 T-weekly-long observations of asset returns and an estimation window of the length $W = 52$.[3] The sample period is from February 2000 until May 2018. The procedure behind the rolling-sample is that in each week t, starting from week $t = W + 1$, the observations from the previous window W are analysed and then utilised to determine the factors, which result in the most optimal portfolio weights that are required to apply to the most profitable asset allocation strategy (in-sample).[4] The determined weights are then utilised in the OOS period, thus they are applied in the week after in order to calculate the portfolio performance, the returns at $t + 1$. Finally, the rolling-sample values are constructed by repeatedly shifting the estimation window one week forward. More precisely, the earliest return of the period is excluded and the next period return included until the last period of the dataset in May 2018 is achieved.

The result of this procedure is the creation of a new sample sequence of T-W weekly OOS returns, initiated by the underlying six portfolio strategies, which are introduced in the next section, and for each empirical dataset (Bessler & Wolff, 2015; DeMiguel et al., 2009; Fethke & Prokopczuk, 2018; Kremer, 2015). The crucial difference between the OOS and the IS is that the latter one uses its asset return estimates based upon the realised returns of that week whereas the OOS uses estimates based upon the next week (Fethke & Prokopczuk, 2018).

In addition, as robustness test relative transaction costs of 7.5 basis points are deducted from every T-W weekly return of the total weekly transaction volume. Short-selling constraints and different economic cycles are also examined, next to the varying window lengths in order to generate more realistic and robust real-world results (Bessler & Wolff, 2015; DeMiguel et al., 2009; Fethke & Prokopczuk, 2018)[5].

[3] For robustness, the results of the OOS strategies with estimation windows of 26 and 104 weeks are reported in section 5.4.1

[4] Portfolio weights are constructed at every trading Wednesday of each week ***t***. More details about this choice are explained in the data section in Chapter 4.2

[5] 7.5 Basis points are the weekly ratio of 30bp per month. A detailed description of the sub-samples is shown in the data section 4.3.

3.2.2 Multiple asset allocation strategies

The generated IS and OOS return samples are selectively tested for maximum six asset allocation strategies. These include the traditional allocation strategies such as (1) the strategically-weighted (SW) and (2) the equally weighted (EW) portfolio construction. Next to that, the portfolio optimisation strategies (3) the maximum Sharpe ratio (MS) and the (4) minimum variance (with rebalancing) (minVar) are conducted. Finally, in order to generate more robust results, (5) naïve risk-parity (RP) and (6) reward-to-risk timing (RRT) strategies are also conducted (Bessler & Wolff, 2015; Fethke & Prokopczuk, 2018; Henriksen, 2018). **Table 1** shows a summary of all strategies.

The simple *1/*N or equally-weighted (EW) asset allocation strategy equally weights all available assets *N* in the portfolio over the whole period. The portfolio weights are simply depicted by $w_i^{(1/N)} = 1/N$. There exist no estimation errors since there are no estimators applied and it does not involve an optimisation approach (Heier Hansen-Tangen & Overaae, 2015).

The second strategy is the strategically-weighted (SW) portfolio, which is similar to the EW method and assets are held constant over time to represent different types of investors and their risk preferences (Bessler & Wolff, 2015). In the first approach, the weights of the commodity indices are strategically allocated to be 5%, 10% and 15% and the remaining weights equally-weighted to the other assets. The second approach furthermore differentiates between aggressive and conservative investors, in which some parts of the contribution to stocks substitute to commodity indices. Initially, the weights of the aggressive investor without commodities contributes 80% to stocks and 20% to bonds compared to the conservative investor, who invests 20% to stocks and 80% to bonds. Including commodities, the aggressive investor now distributes 65% to stocks and 15% to commodities and the conservative investor 15% to stocks and 5% to commodities. The weights to bonds stay the same for both types ((Bessler & Wolff, 2015; Conover, Jensen, Johnson, & Mercer, 2010; Erb & Harvey, 2016). In general, naïve diversification strategies are more applied by private investors (Benartzi & Thaler, 2001).

Table 1: Summary of the asset allocation strategies

This table presents a summary of the asset allocation strategies applied in this study. The names in the abbreviation column refer to the names in the output tables, which display the empirical results. The Application column indicates in which analysis the strategies are employed.

#	Strategy	Abbreviation	Application
Naïve allocation strategies			
1	Equally-weighted	EW	IS
2	Strategically-weighted	SW	IS
Simple allocation strategies			
3	Risk-Parity	RP	IS & OOS
4	Reward-to-risk timing	R2R	IS & OOS
Portfolio optimisation strategies			
5	Minimum Variance	minVar	IS & OOS
6	Maximum Sharpe ratio	maxSR	IS

Therefore, more simple asset allocation strategies, like (3) the naïve risk-parity (RP) strategy and (4) the reward-to-risk timing (R2R) strategy are applied that do not involve optimisation and in order to generate robust results. Whereas the former one is increasingly applied by long term investors like pension funds, or index providers and became more important after the financial crisis in 2008 (Anderson, Bianchi, & Goldberg, 2012), the latter one is more interesting for risky investors because of its fundamental strategy, which considers expected asset returns (Bessler & Wolff, 2015).

The idea of the RP is that each asset class contributes equally to the portfolio risk. Thus, it equalises the share of every asset class on the overall portfolio risk (Asness, Frazzini, & Pedersen, 2012). In this thesis, Bessler and Wolff's (2015) approach is used because it weights the assets anti-proportional to its sample variance and does not account for asset return correlation.

$$\omega_i^{RP} = \frac{1/\hat{\sigma}_i^2}{\sum_{i=1}^{N}(1/\hat{\sigma}_i^2)} \tag{8}$$

The RP is predominantly used by long-term investors such as pension funds since it makes use of the low-volatility anomaly and usually performs better than the naïve portfolio models (Bessler & Wolff, 2015).

The last strategy is the reward-to-risk timing (R2R) strategy that determines the weights of the assets according to the reward-to-risk ratio of the specific assets and is based on Kirby and Ostdiek's (2012) approach. This ratio calculates the sample mean of the assets over the sample variance as follows

$$\omega_i^{R2R} = \frac{\hat{\mu}_i^+/\hat{\sigma}_i^2}{\sum_{i=1}^{N}(\hat{\mu}_i^+/\hat{\sigma}_i^2)} \tag{9}$$

where $\hat{\mu}_i^+$ = max ($\hat{\mu}_i$, 0). The RRT strategy tries to lessen extreme allocations, does not enable short-selling and eliminates samples with negative expected returns, by defining $\hat{\mu}^+$ and neglecting $\hat{\sum}$. If all expected asset returns are negative so that Equation (9) cannot determine the weights, the wealth is equally distributed among all available assets (Fethke & Prokopczuk, 2018), which is the case for the last year observations in 2018.

The third strategy is the maximum Sharpe (maxSR) portfolio. It is also called the tangency portfolio and shows the location where the asset allocation line is tangent to the efficient frontier. The formula is

$$Max_w\ Sharpe\ ratio = \frac{\overline{r_p} - r_f}{\sigma_p} \quad s.t. \quad \sum_{t=1}^{n} w_i = 1 \text{ and } 0 \leq w_i \leq 1 \tag{10}$$

where r_p is the average return of the portfolio, r_f is the average risk-free rate and σ_p the standard deviation, so the volatility of the portfolio.

The results of the strategies are expected to yield different SRs. Whereas the minVar and the RP try to invest in low volatility assets to diversify portfolio risk, the maxSR and the R2R are rather risk-seeking strategies to maximise their reward-to-risk ratio. The results of the different strategies indicate well the lower SRs for the safer and higher SRs for the risky ones.

Finally, each asset allocation strategy is applied to the benchmark and then to the augmented portfolios with the commodity indices. Then the applicable IS and OOS portfolio returns are calculated and compared.

3.2.3 Multiple performance measures

Eventually, the performance enhancement of the augmented benchmark portfolios due to the inclusion of the commodity indices is evaluated by several measures. For both, the IS and well as the OOS results the average mean and volatility of returns are examined and also the net Sharpe ratio (SR), which is the average net excess return divided by the volatility of the net average excess returns. The significant difference of the SR between the initial and the augmented portfolio is checked by the modified approach from Memmel (2003), who improved Jobson and Korkie's (1981) approach. The calculated z-score is then converted to the corresponding p-value using a two-tailed hypothesis testing. Significance levels of the differences between the SR's at the 0.01, 0.05, and 0.10 significance levels, are indicated by the asterisk in the result tables.

Next to that, the average commodity weight is evaluated for IS as well as OOS performances. It is the average portfolio weight of each commodity indices over the *T-M* rebalancing periods for each strategy *i* and given by

$$Average\ commodity\ weight_i = \frac{1}{T-W} \sum_{t=1}^{T-W} \omega_{i,t}^{com} \tag{11}$$

where $\omega_{i,t}^{com}$ is the portfolio weights of commodities at time *t* and for strategy *i*. The calculation of this measure analyses the effect of commodities on the results of each asset allocation strategies and also the optimal relative allocation towards commodities (Fethke & Prokopczuk, 2018).

Furthermore, all OOS investigations are adjusted after transaction costs. In the end, portfolio turnover (PT_i) values for strategy *i* are also calculated according to the approach of DeMiguel et al. (2009). In general, the PT calculates the amount of rebalancing, which is required to implement strategy *i*. It is defined as the average sum of absolute changes in portfolio weights *w* over the *T-K* rebalancing points and across the *N* available assets, i.e.

$$PT_i = \frac{1}{T-W} \sum_{t=1}^{T-W} \sum_{j=1}^{N} (|\omega_{i,j,t+1} - \omega_{i,j,t+}|) \tag{12}$$

where $|\omega_{i,j,t+1} - \omega_{i,j,t+}|$ are the optimal weights of asset j under strategy *i* at time t and t+1, respectively. W_{t+} is the portfolio weight prior to the rebalancing point and wt+1 is the weight after the rebalancing at t+1 (Fethke & Prokopczuk, 2018). The resulting quantity of $|\omega_{i,j,t+1} - \omega_{i,j,t+}|$ equals the trading volume of asset j to fill the gap at the rebalancing point t+1. Stated differently, it is the proportion of portfolio value (in percentage terms) that needs to be reallocated over the complete period (Daskalaki et al., 2017).

CHAPTER 4

Data

This chapter describes the separate generation commodity, stock and bond indices, which are selected in order to create traditional portfolios and to statistically analyse the impact of commodities on these.

4.1 Commodity indices

In order to represent the current commodity market appropriately, a variety of diverse commodity generations is considered. The choice of indices reflects a smaller sample of Fethke's and Prokopczuk's (2018) selection. Furthermore, since their results reveal diverging performances of 3rd generation indices, this study selects indices from both groups in order to challenge their findings and controls it by also checking for 1st and 2nd generation indices.

As can be seen in **Table 2**, there are in total 7 different indices and the main focus lies on 3rd generation indices. Four out of the seven indices are 3rd generations and the complementary ones include two 2nd and one 1st generation indice. Following, a brief outline of every index is provided next to the reasons for the inclusion. A more detailed description of the indices can be found in **Table 12** in the Appendix.
For the 1st generation commodity indices, the most widely accepted Standard & Poor's Goldman Sachs Commodity Index (SPGSCI) is employed. The SPGSCI is the most popular first generation index within the commodity indexes worldwide (Skiadopoulos, 2012) and captures the market premium of long holdings in commodity derivatives positions. It aims at covering a broad range of commodities, even though it is biased since it dedicates large weights (about 70%) towards the energy sector of the total index value.
For the **2nd** generation indices the Deutsche Bank Liquid Commodity Index Optimum Yield Balanced (DBOYBA) and the SummerHaven Dynamic Commodity Index (SDCI) are employed. Both indices are the best performing 2nd generations in Fethke's and Prokopczuk's (2018) paper and are employed in order to show OOS outperformances and to challenge the 3rd generation indices. The SDCI looks at fundamental signals about the underlying physical markets to build an active commodity portfolio and rebalances monthly.[6] The DBOYBA invests in 14 long positions based on the most backwardated future contracts (Fethke & Prokopczuk, 2018).
The **3rd** generation commodity indices analysed in this thesis are the Morningstar Short-Only Commodity Index (MSSO), the Credit Suisse Momentum and Volatility Enhanced Return Strategy (CSMOVERS), the CYD Market Neutral Plus Commodity Index (CYDMNP) and lastly, the Barclays Backwardation Long/Short Index (BBLS). According to Fethke & Prokopczuk's (2018) results, all 3rd

[6] for more information, see: https://summerhavenindex.com/sdci/

generation indices, except for the MSSO, belong to their outperforming group. BBLS is added in this study because it is characterised by going long and short in six commodities depending on the term structure of future contracts. More information can be found in **Table 12**.

Finally, all the indexes consist of robust and rich data sets requisite for the analysis. Next to that, this selection decision tries to incorporate a wide range of different investment strategies and index providers.

Table 2 List of asset classes

Notes: This table summarises all asset classes applied in this study. 1st, 2nd and 3rd Generation refer to the commodity index generations. Stocks & Bonds refers to all other traditional asset classes applied to create the portfolio. Asset class 8 to 11 present the stock indices, 12 to 15 the bond indices. 16 refers to the S&P500 index and 17 is the risk-free rate. All indices are denoted in USD. ICE BofAML stands for Bank of America Merrill Lynch.

Asset class #	Datastream or Bloomberg Code	Abbreviation	Name
1st Generation			
1	GSCITOT(TR)	GSCI	S&P Goldman Sachs Commodity Index
2nd Generation			
2	DBLCBBTR	DBOYBA	Deutsche Bank Liquid Commodity Index Optimum Yield - Balanced
3	SDCITR	SDCI	SummerHaven Dynamic Commodity Index
3rd Generation			
4	MSDISTR	MSSO	Morningstar Short-Only Commodity Index
5	CSMVRSTR	CSMOVERS	Credit Suisse Momentum & Volatility Enhanced Return Strategy Index
6	CYDIMNTR	CYDMNP	CYD Market Neutral Plus Commodity Index
7	BCCFBKAT	BBLS	Barclays Backwardation Long Short Index
Traditional portfolio assets			
8	MSUSAML(MSRI)	MSUS	MSCI USA - Total Return Index
9	MSEROP$(MSRI)	MSEU	MSCI EUROPE - Total Return Index
10	MSEMKF$(MSRI)	MSEM	MSCI EM - Total Return Index
11	M2DWR2$(MSRI)	MSRE	MSCI World Real Estate - Total Return Index
12	MLGCOR$(RI)	MLGC	ICE BofAML Global Corporate Index - Total Return Index
13	MLTRSML(RI)	MLTR	ICE BofAML US Treasury Index - Total Return Index
14	MLGGIL$(RI)	MLGG	ICE BofAML Global Inflation-Linked Government Index - Total Return Index
15	JPEIDIVR Index	JPEI	JP Morgan - EMBI Global Diversified Composite
S&P 500 and risk-free rate			
16	SPX Index	SP500	S&P 500 Index
17	FRTBS3M(RI)		US T-Bill Secondary Market 3 month (daily)

4.2 Data selection

According to Henriksen's approach (2018), and as can be seen in **Table 2** this thesis uses a broader investment universe to create the benchmark portfolio. All assets are denominated in U.S. dollar and the diversified portfolio consists of in total four stocks and four bond indices. The MSCI U.S., Europe and the Emerging Markets are used to represent the global stock market. The MSCI Real Estate index complements this asset class in order to create a mixed-asset portfolio that might potentially improve the risk-return profile (NBIM, 2015). In order to provide more diversification effects, four bond indices,

such as the corporate, Treasury, and government inflation-linked bond indices provided by Bank of America Merrill Lynch (BofAML) are included. Next to that, the JP Morgan global diversified emerging market index is used to include the bond market of emerging countries. In this thesis, long only stock and bond indices are used since they result in risk premiums over time (Henriksen, 2018). The return of the weekly U.S. Treasury Bill is used as risk-free rate and the complete data set is denominated in US dollar and investigates the period from February 1st, 2000 until May 31st, 2018 in weekly frequencies of Wednesday prices[7]. Weekly measures are examined because they are more appropriate to account for historical tail-risks than lower frequencies (Henriksen, 2018).

In general, for the analysis, time series data points are indexed in total returns for all strategies during all IS and OOS analyses. For the spanning tests, simple excess returns are used, which are the total net return subtracted by the daily risk-free lending rate (Belousova & Dorfleitner, 2012; Kan & Zhou, 2012; Yan & Garcia, 2017).

One contribution to the literature is the distinction between the period before the launch and after the launch of the last 3rd generation commodity indices, the BBLO index (later referred to as before-launch and after-launch periods). Consequently, the analyses compare the period before the launch from February 2000 until November 2010 and especially after the launch from December 2010 until May 2018. Even though some indices were launched already earlier than December 2010, the analysis of the period after the launch is decided to be the date of the launch of the BBLO to incorporate only real performances of all indices (Fethke & Prokopczuk, 2018).

Besides, the dataset is divided into bullish and bearish market periods as can be seen in **Figure 4** and **Table *3*** in order to compare the effects on varying bull and bear market periods (Belousova & Dorfleitner, 2012). **Figure 4** also shows the prices of the S&P 500 stock market index as a curve to visually distinguish the periods that are equal to the diverging phases of the business cycles. In the end, for the spanning tests, four different sub-periods are chosen to evaluate the performance of commodities during these market trends, since the selected periods depict alternating bearish and bullish cycles.

In the end, there are in total 956 time-series data points over the complete time period but fewer for the sub-periods, as indicated in

Table *3*. Stock, bond, commodity and risk-free rates data are mainly retrieved from Thomson Reuters DataStream. The completing data come from Bloomberg and the index provider website of SummerHaven[8].

[7] "Mid-weekday prices are used in order to eliminate the potential weekend effect. In case of a non-trading Wednesday, the previous or closest trading day available is taken" (Henriksen, 2018).

[8] https://summerhavenindex.com/

Table 3: List of sub-periods

Notes: This table lists the beginning and ending dates of all examined sub-periods and also the number of observations per period.

Time period	Starting	Ending	Observations
Complete period	02.2000	05.2018	956
1st period	02.2000	04.2003	169
2nd period	05.2003	10.2007	235
3rd period	11.2007	03.2009	73
4th period	04.2009	05.2018	479
Before launch period	02.2000	11.2010	564
After launch period	12.2010	05.2018	392

CHAPTER 5

IS and OOS results

In this chapter the descriptive statistics and the results of the spanning, IS, and OOS calculations are presented.

5.1 Descriptive Statistics

As a short overview of the different asset classes, this section depicts the basic descriptive statistics.
In

Table *4* the average return over the complete period yields different results. The GSCI, as the only example for 1st generation indices, shows the second lowest annualized return just above the average risk-free rate of 1.59% and the highest volatility with 22.4%. Consequently, the SR is also the second lowest of all observations and underperforms the S&P 500 considerably.
The 2nd generation indices show higher returns and lower volatilities than the GSCI and therefore higher SR's than the S&P 500 index. Their risk-return profile hence hints to be good investment opportunities.
The 3rd generation indices show ambivalent results but most notably are the BBLS and the CYDMNP indices that show the highest mean return or the lowest volatility from the complete sample, respectively. As Fethke and Prokopczuk (2018) already discovered and in contrast to the BBLS, the CYDMNP outperforms the other indices due to its specific investment strategy and its leverage factor. The leverage however should not affect the SR to a great extent, so they conclude that the strategy influences the performance the most. In the end, the two indices experience the highest SR with 1.04 and 1.47 of the whole sample, respectively. On the contrary, the MSSO presents the worst performance of all indices examined in this study. An explanations for this result is the strategy that the index does not invest in long positions (Fethke & Prokopczuk, 2018). Furthermore, the CSMOVERS shows positive performances that are similar to the 2nd generations.
In conclusion, the 1st generation index does not seem to be a good stand-alone investment, whereas 2nd and 3rd generations in general seem to add value and even outperform the S&P 500 index in terms of the volatility and thus SR. Moreover, 3rd generation indices show mixed results and whereas some show promising performances, others do not or even negative results.

The traditional assets such as the stock indices and the S&P500 index show mediocre returns between 6% and 9% and a relatively high volatility between 16.53% and 21.68%. On contrary, the bond indices exhibit on average the lowest volatilities between 4% and 7% and means between 4% and 9%. Hence, the JPEI shows the third highest SR of all indices with 1.01.

Table 4 Descriptive Statistics

Notes: This table reports the descriptive statistics for all asset classes over the period from February 2000 until May 2018. Calculations are based on weekly total return data. Mean and standard deviation (StDef) refer to the annualized percentage values. T-stat reports the t-statistics of a standard t-test with the null hypothesis of an annualized mean equal to zero. The Sharpe ratio equals the annualized time series mean of excess returns divided by the annualized volatility of weekly net returns. Ex. Kurt. refers to the excess kurtosis and the Jarque-Bera (JB) p-values test for normality of returns and the null hypothesis indicates that the calculated asset returns follow a normal distribution. The average risk-free rate during 2000-2018 is 1.59%. VaR and ES are the risk measures value-at-risk and expected shortfall, respectively.

Index	Mean	t-stat	StDev	Sharpe	Ex. Kurt.	Skewness	JB	VaR (5%)	ES (5%)
SPGSCI	2.20	0.42	22.38%	**0.03**	1.25	-0.19	67.9	-34.64%	-6.97%
DBOYBA	8.16	2.12	15.84	**0.41**	2.08	-0.36	192.9	-18.21%	-5.23%
SDCI	11.24	3.09	14.79	**0.64**	2.10	-0.50	216.4	-13.66%	-4.87%
MSSO	-2.10	-0.53	17.05	**-0.21**	1.11	0.17	53.5	-30.17%	-4.88%
CSMOVERS	9.43	2.60	14.87	**0.52**	3.48	0.40	508.5	-15.43%	-4.13%
CYDMNP	5.93	8.35	2.96	**1.47**	9.46	-0.28	3576.7	0.90%	-0.79%
BBLS	12.58	4.89	10.41	**1.04**	1.16	-0.13	56.4	-5.26%	-2.96%
MSUS	7.28	1.76	16.53	**0.32**	5.22	-0.42	1113.2	-20.40%	-5.52%
MSEU	7.02	1.35	20.42	**0.24**	2.99	-0.48	393.2	-27.17%	-6.91%
MSEM	6.63	1.87	21.68	**0.38**	6.02	-0.74	1529.1	-26.19%	-7.17%
MSRE	9.92	2.04	18.92	**0.41**	6.52	-0.72	1774.7	-22.11%	-6.61%
MLGC	9.42	4.28	5.36	**0.72**	4.66	0.13	869.0	-3.47%	-1.59%
MLTR	5.49	4.43	4.39	**0.69**	0.73	-0.09	22.7	-2.69%	-1.24%
MLGG	4.64	3.47	7.20	**0.60**	5.84	0.30	1374.2	-6.02%	-2.11%
JPEI	6.00	5.10	7.44	**1.01**	20.66	-1.20	17229.5	-3.38%	-2.46%
SP500	9.24	1.28	16.61	**0.21**	4.66	-0.43	895.2	-22.35%	-5.54%

All asset classes show excess kurtosis values and hence have fatter tails than under a normal distribution. The high JB values further endorses this conclusion that the assets are non-normally distributed. Only the MSSO, CSMOVERS, MLGC, and MLGG show positively skewed returns, thus they have rather upside risks. The rest are rather negatively skewed and have more downside risks.

Moreover, all equities show similar high VaR values compared to the bonds and CYDMNP and BBLS. The latter ones also have lower expected shortfall values.

In summary, all commodity indices and especially 3rd generation indices, except for the MSSO, yield positive returns and SR's, and outperformed the risk-free rate of 1.59% on average over the complete period. For the rest of the paper, average risk-free rates of 2.56% and 0.32% are used to calculate the SR for the before launch and after launch period, respectively.

Besides the descriptive statistics, the correlation of the commodities with the traditional assets is another appropriate indicator to assess the potential diversification benefits of the alternative asset class. As can be seen in the correlation matrix in **Table 5**, 1st and 2nd generation commodities are generally highly correlated with each other but not correlated or negatively correlated with the 3rd generation. Furthermore, they show low correlation with stocks and bonds in a range of 0.27 and 0.47, and only the correlation with the MLTR is slightly negative. These results might imply that there exist

Table 5: ***Correlation matrix of all asset classes***

Notes: This table shows the Pearson's correlation matrix of all employed asset classes over the complete period from 02.2000 until 05.2018. The *, **, *** indicate if significance values are different to 0 at the 10%, 5%, and 1% significance level, respectively.

	GSCI	DBOYBA	SDCI	MSSO	CSMOVERS	CYDMNP	BBLS	MSUS	MSEU	MSEM	MSRE	MLGC	MLTR	MLGG	JPEI	SP500	Rf
BOYBA	0.89***	1															
SDCI	0.79***	0.91***	1														
MSSO	-0.92***	-0.91***	-0.82***	1													
CSMOVERS	0	0.04	0.13***	0.06*	1												
CYDMNP	-0.34***	-0.17***	-0.17***	0.31***	-0.07**	1											
BBLS	0.08**	0.10***	0.25***	0.07**	0.35***	0.08***	1										
MSUS	0.30***	0.34***	0.35***	-0.33***	-0.16***	0.02	0	1									
MSEU	0.37***	0.44***	0.44***	-0.40***	-0.12***	0.02	0.01	0.77***	1								
MSEM	0.35***	0.45***	0.47***	-0.39***	-0.09***	0.08**	0.07**	0.67*	0.79*	1							
MSRE	0.27***	0.35***	0.37***	-0.31***	-0.13***	0.06*	0.02	0.74***	0.74***	0.76***	1						
MLGC	0.19***	0.29***	0.27***	-0.22***	-0.04	0.01	0.03	0.08***	0.29***	0.27***	0.31***	1					
MLTR	-0.1872***	-0.16***	-0.17***	0.19***	0.04	0.03	0.01	-0.32***	-0.31***	-0.25***	-0.14***	0.56***	1				
MLGG	0.24***	0.32***	0.28***	-0.28***	-0.02	0	0.01	0.06*	0.26***	0.22***	0.26***	0.87***	0.52***	1			
JPEI	0.2470***	0.31***	0.30***	-0.26***	-0.15***	0.04	0.08**	0.36***	0.44***	0.57***	0.50***	0.49***	0.16***	0.41***	1		
SP500	0.29***	0.33***	0.33***	-0.32***	0.33***	0.02	0	0.94***	0.75***	0.64***	0.70***	0.07**	-0.33***	0.05*	0.39***	1	
Rf	0.04	0.06*	0.08**	0.00	0.09***	0.19***	0.11***	-0.05	-0.02	-0.02	-0.01	0	0.07**	0.01	0	-0.04	1

some potentially but expected diversification benefits of 1st and 2nd generations with traditional assets. Next to that, investment in these two generations show similar behaviours, since both cases invest in long-only future contracts. In comparison to that, the 3rd generation commodities also invest in short positions and hence their correlation coefficient to 1st and 2nd generations is non-uniform and shows highly negative numbers for the MSSO and CYDMNP ranging from -0.92 to -0.17, or slightly positive results for the CSMOVERS and BBLS ranging from 0.08 to 0.25. Next to that, the correlation to the traditional asset classes are in general negative for the MSSO and the CSMOVERS. Slightly positive of about 0.08 or no significant correlation are shown for the CYDMNP and the BBLS.

One explanation for the deviating observation is the sophisticated investment strategy of 3rd generations, which includes short positions (Fethke & Prokopczuk, 2018).

Table 6: Descriptive statistics before and after launch of all assets

Notes: This table reports the descriptive statistics for all asset classes for the periods before and after the last launch in December 2010. The period before the launch includes 564 observations and the after launch 392 observations. The average risk-free rate during 2000-2010 is 2.51% and from 2011-2018 is 0.28%. Entries are based on weekly total return data. Mean and StDev refer to the annualised return and standard deviation in percentage values. Sharpe refers to the Sharpe ratio, which is calculated as annualized time series mean of excess returns divided by the annualized volatility of net excess returns. Ex.Kurt refers to the excess kurtosis and JB to the Jarque-Bera p-values that test for normality of returns and the null hypothesis states that asset returns follow a normal distribution.

	Mean	StDev	Sharpe	Ex. Kurt	Skewness	JB	VaR
Before launch							
GSCI	7.21	24.92	**0.19**	1.01	-0.31	33.00	33.79
DBOYBA	16.02	17.45	**0.77**	2.02	-0.51	120.48	12.68
SDCI	20.68	16.49	**1.10**	1.83	-0.69	122.83	6.45
MSSO	-5.19	18.29	**-0.42**	1.05	0.32	35.72	35.28
CSMOVERS	18.58	16.12	**1.00**	3.65	0.41	328.91	7.94
CYDMNP	9.15	3.43	**1.93**	8.33	-0.52	1657.86	-3.50
BBLS	19.32	11.35	**1.48**	0.87	-0.13	19.37	-0.64
MSUS	1.92	18.57	**-0.03**	4.37	-0.25	454.95	28.63
MSEU	5.57	21.96	**0.14**	2.75	-0.56	208.31	30.55
MSEM	13.81	24.09	**0.47**	6.17	-0.86	965.51	25.82
MSRE	9.55	21.84	**0.32**	5.45	-0.67	741.38	26.38
MLGC	7.00	5.96	**0.75**	4.63	0.21	508.86	2.80
MLTR	6.38	4.81	**0.80**	0.61	-0.19	12.22	1.53
MLGG	7.83	7.75	**0.69**	6.90	0.46	1138.13	4.93
JPEI	11.83	8.43	**1.11**	20.68	-1.29	10207.8	2.04
After launch							
GSCI	-4.61	18.10	**-0.27**	0.74	0.13	9.99	34.38
DBOYBA	-2.24	13.09	**-0.19**	1.22	-0.07	24.70	23.78
SDCI	-1.09	11.73	**-0.12**	2.12	-0.12	74.00	20.39
MSSO	2.52	15.09	**0.15**	0.92	-0.16	15.48	22.29
CSMOVERS	-2.53	12.70	**-0.22**	1.41	0.14	33.78	23.42
CYDMNP	1.46	1.95	**0.61**	3.26	-0.25	177.55	1.74
BBLS	3.53	8.77	**0.37**	1.29	-0.41	38.61	10.88
MSUS	14.80	13.02	**1.12**	5.43	-0.86	530.39	6.61
MSEU	8.17	18.01	**0.44**	2.95	-0.22	145.44	21.46
MSEM	4.56	17.64	**0.24**	1.75	-0.28	55.09	24.46
MSRE	9.22	13.68	**0.65**	3.24	-0.70	203.86	13.27
MLGC	3.36	4.35	**0.71**	1.33	-0.35	37.18	3.79
MLTR	2.19	3.70	**0.52**	0.43	0.05	3.26	3.90
MLGG	3.42	6.32	**0.50**	1.04	-0.23	21.04	6.98
JPEI	5.61	5.68	**0.94**	2.83	-0.76	167.96	3.73

In summary, according to the correlation matrix, 3rd generation commodities exhibit superior characteristics for diversification to the previous generations but deviate significantly within their own group.

Concluding, similar to Fethke and Prokopczuk (2018), 1st, 2nd, and 3rd generations commodity indices exhibit attractive diversification opportunities due to the low correlation coefficients to traditional assets. Furthermore, even though 1st and 2nd generations have highly correlated returns, 2nd generations show better results and outperform the 1st generations in terms of higher returns and a lower volatility. Additionally, 3rd generations are not correlated to 1st and 2nd generations and also not to traditional assets and hence indicate better diversification characteristics. Nevertheless, they also show mixed results in terms of the return, volatility and the SR and therefore it is not appropriate to generalise this generation due to performance differences. These first findings are in line with previous evidence shown above.

Since the fact that most of the generation commodity indices usually show a large amount of backfilled data points and most researcher did not consider this to a great extent, this study examines the performance of commodity indices differentiated for the before launch and after launch period, as describe in the data section 4.2. **some potentially** but expected diversification benefits of 1st and 2nd generations with traditional assets. Next to that, investment in these two generations show similar behaviours, since both cases invest in long-only future contracts. In comparison to that, the 3rd generation commodities also invest in short positions and hence their correlation coefficient to 1st and 2nd generations is non-uniform and shows highly negative numbers for the MSSO and CYDMNP ranging from -0.92 to -0.17, or slightly positive results for the CSMOVERS and BBLS ranging from 0.08 to 0.25. Next to that, the correlation to the traditional asset classes are in general negative for the MSSO and the CSMOVERS. Slightly positive of about 0.08 or no significant correlation are shown for the CYDMNP and the BBLS.

One explanation for the deviating observation is the sophisticated investment strategy of 3rd generations, which includes short positions (Fethke & Prokopczuk, 2018).

Table *6* presents the descriptive statistics of all assets for the periods before and after the last commodity index launch in December 2010. As a side-fact, this period experiences one of the longest bullish market trends, which is represented in the scores of the assets, as well. Since index provider release more data points based on backfilled calculated information, the returns of the time before the launch should be evaluated with attention. Therefore, in this thesis the focus is on the returns of the period after the launch that present only the actual performance without backfilled data observations.

In general, all assets show lower SRs in the after launch period than before the launch, which is depicted in lower volatility as well as return levels, expect for the MSSO, which is the only improving index. Another interesting finding is that only the 3rd generation indices MSSO, the CYDNP and the BBLS show positive SRs as well as all the other traditional assets. The CSMOVERS and the 1st and 2nd generation indices do not show positive risk-return profiles for the after launch period anymore. Next to that, commodities show lower volatility scores than stocks but not than bonds. In contrast to that, bonds, except for the MLTR, show higher returns and stocks even substantially higher scores than all commodity indices.

Concluding, the backfilled observations of 2nd as well as 3rd commodity indices show better results than the actual performance of the indices after the launch. Accordingly, the former period should be evaluated more carefully to analyse the usefulness of commodity indices to a traditional portfolio. Finally, this thesis emphasises on the period after December 2010 in order to evaluate the actual performance commodity indices during the bullish period. According to the descriptive statistics, no 1st and 2nd generation indice and only some 3rd generation indices suggest to be useful for investors after the launch, however more analyses need to be performed to empirically prove these assumptions.

5.2 In-sample results

5.2.1 IS mean-variance spanning test results

In

Table *7* the results of the spanning and step-down tests without rebalancing for the complete time period are listed. The spanning tests analyse if there is a statistical improvement of the efficient frontier between the benchmark and the commodity indices enhanced portfolio. The step-down tests indicate if the source of improvement is due to higher returns or due to lower volatility values. The null hypothesis states that there is no significance difference between the benchmark and the augmented portfolio.

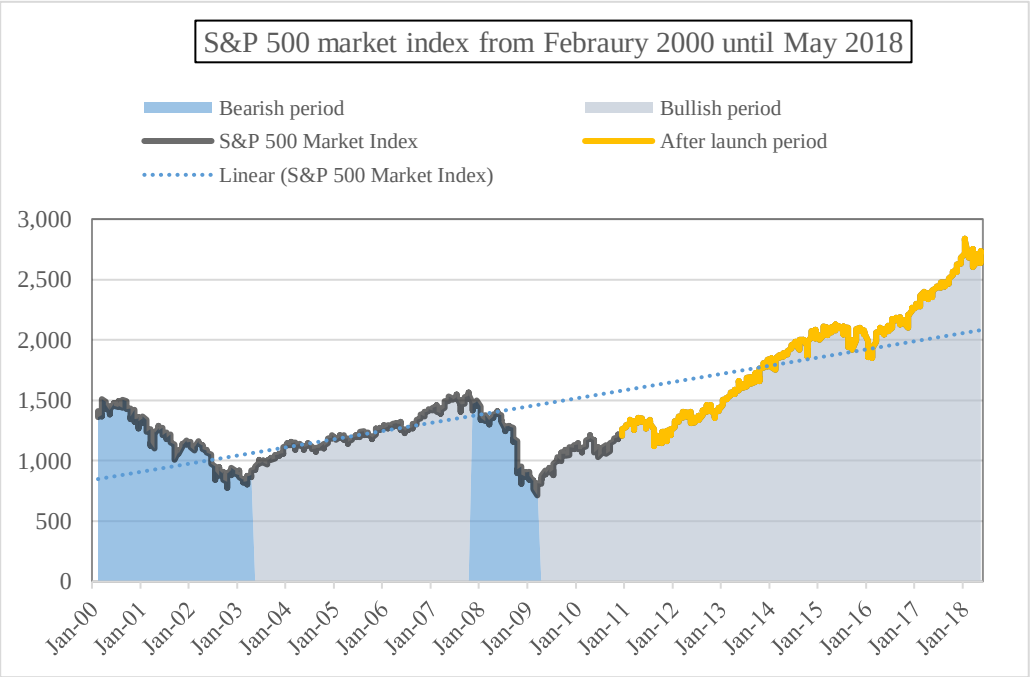
As can be seen in

Table *7*, all indices are significant and can reject the null hypothesis of spanning, so there are strong evidences for diversification effects of commodities over the complete period (Belousova & Dorfleitner, 2012). Therefore, all indices are able to improve the benchmark portfolio at the 1% significance level,

even when non-normality and heteroscedasticity is considered by the GMM-Wald test. Besides, 3rd generation indices show the highest coefficients, indicating the strongest influence on the performance. The CYDMNP for instance shows for all tests significantly higher values than the other indices. The results of the step-down tests show mixed results. The F_1 test for instance is not significant for the GSCI, DBOYBA and the MSSO. All the other indices however show significant results at the 1% level. Next to that, the F_2 test is significant for all commodity indices and indicates high values, meaning that commodity indices rather lower the risk than increasing the return of a portfolio.

Concluding, all indices are able to add value to a diversified portfolio over the complete period. Furthermore, their usefulness is due to their low correlation and thus the performance improvement is rather driven by lowering the overall volatility of a portfolio than by improving the return. Additionally, 3rd generation indices on average depict stronger effects than previous generations but they also show different performances within that generation. The overall conclusion is in line with Fethke & Prokopczuk (2018), who however do not show significant results for 2nd generation indices. Differences can be explained due to a longer time period observed and the more sophisticated portfolio in this study. Apart from that, given the indices descriptive statistics', the correlation matrix and the findings of Kremer (2015), the results are justified because of their individual risk-return profile and the low correlation to the traditional asset classes.

Figure *4*: S&P 500 Market Index



This figure shows the S&P 500 market index from February 2000 until May 2018. Furthermore, it marks the bearish and bullish periods in light blue or light grey colours. The yellow line indicates the index for the period after the launch of the last commodity index in December 2010. The blurred line indicates the market trend.

Table *13* in Appendix B shows similar values for the period before the last launch in December 2010. The added value of generation commodity indices is again mainly driven by the risk reduction and to a lesser extent by a return advancement as indicated by the higher F_2 and lower F_1test value, respectively. The evidence for the tests after the launch in **Table 14** however show contrary results. Surprisingly, all indices, except for the GSCI at the 10% level, do not show to improve the return of a portfolio after December 2010 according to the F_1 test. These results are in line with the previous findings since the market trend after the launch period is a bullish market period and all generation commodity indices did not demonstrate to yield better return performances than the traditional asset classes. Nevertheless, the F_2 tests are significant again and suggest that each index is still able to reduce the overall portfolio risks. In summary, commodity indices do not seem to improve portfolio returns for the after launch period but are able to do so for the before launch period. Questionable in this case is if the different market trend plays a role in the performance and if the indices increase returns rather during bearish or bullish periods, since the after launch period depicts a very long bullish trend. Nonetheless, all indices are able to reduce the volatility for all periods, which can be attributed to the low return correlations.

Finally, the outcomes of the mv spanning and step-down tests demonstrate that that the inclusion of commodity indices in general can significantly add value to an already diversified portfolio during all times (Belousova & Dorfleitner, 2012). Furthermore, the results propose that the improvement is primarily due to risk reductions, as indicated by significant and higher F_2 test values, but to a lesser extend due to the return improvement, because of lower and more insignificant F_1 test values. Depending on the market trend however, 2nd and 3rd generations are potentially also able to enhance portfolio returns. Apart from that, the actual performance of all commodity indices after their launch is different than the performance of their backfilled data period as suggested in the descriptive statistics section 5.1. Therefore, it is necessary to differentiate between live-data and backfilled data, as well as between bullish and bearish market trends, since the after launch period is predominantly characterised by a very long bullish market period. In the end, low correlations of alternative to traditional asset classes expect to yield suitable diversification features during predominantly bearish periods.

Table 7: Spanning test results over the complete period from February 2000 until May

Notes: This table presents the test statistics and p-values (in parentheses) of the spanning and step-down test results for all commodity indices over the complete period. The index column specifies the commodity index that is examined against the benchmark portfolio consisting of the 4 bonds and stocks. The columns Wald, LR, and LM and GMM-Wald refer to the Wald, Likelihood ratio, Lagrange multiplier and GMM-based Wald test. F_1 and F_2 present the step-down procedure and indicate if the source of performance enhancement is either due to return improvements or risk reductions, respectively. The examination is based on weekly excess returns over the 3-month US T-Bill risk-free rate. The observations cover the appropriate period as indicated in

Table *3*.

	Wald	LR	LM	GMM-Wald	F_1	F_2
GSCI	40.06	39.25	38.45	29.2	0.18	39.55
	(0.00)	(0.00)	(0.00)	(0.00)	(0.68)	(0.00)
DBOYBA	46.91	45.80	44.72	28.34	1.21	45.25
	(0.00)	(0.00)	(0.00)	(0.00)	(0.27)	(0.00)
SDCI	61.65	59.74	57.92	36.48	5.07	55.76
	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)
MSSO	54.69	53.18	51.73	39.96	0.21	54.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.65)	(0.00)
CSMOVERS	95.74	91.24	87.03	67.06	7.75	86.47
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
CYDMNP	1347.40	840.68	559.22	1483.50	37.83	1248.40
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BBLS	103.77	98.52	93.61	105.57	15.86	85.59
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

5.2.2 IS results of equally and strategically weighted strategies

In **Table 8**, the IS equally and strategically weighted results are displayed over the complete time period. Starting with the EW results, all augmented portfolios show significantly higher SR's than the benchmark portfolio, except for the 1st generation index. Next to that, all extended portfolios yield higher means and lower volatility, except for the MSSO, which shows lower returns but also the lowest standard deviation as shown in the descriptive statistics.

The results of the SW analyses indicate similar outcomes and except for the GSCI and in the case with 15% in commodities for the MSSO, all the other indices yield better SR's and higher returns and lower standard deviations. The output for the aggressive and conservative SW investor further support these findings. On average, the 3rd generation indices show greater outperformance results than 2nd generations, however there are diverse outputs within 3rd generation indices again, which show worse ones compared to the 2nd generations. Accordingly, one exception in some cases is the MSSO, which underperforms the benchmark portfolio's mean but also exhibits lower volatility coefficients. These findings are in line with the previous findings and Fethke and Prokopczuk (2018), who show ambivalent performances of 3rd generations, as well. Next to that, the conservative investor, who invest 5% in commodities depicts the highest portfolio SR's of all static strategies with a maximum of 0.98 with the BBLS and due to the low risk-profile. Compared to that, the aggressive investor underperforms and experiences the lowest SR's with a maximum of 0.56 with the BBLS.

Table 8: Naïve diversification results over the complete period

Notes: This table reports the naïve diversification strategies results over the complete observation period from 02.2000 until 05.2018. In particular, it reports the one EW and all SW strategy results. Benchmark refers to the diversified portfolio without commodities. 5%, 10%, 15% refer to the percentage contribution towards

commodities. *Aggr*. refers to a portfolio with an aggressive SW investor, *Cons*. refers to a conservative SW investor. Benchmark indicates the results of the traditional portfolio and the indices abbreviations refer to the augmented portfolios including the commodity index. Mean, StDev, Com weight, and Sharpe stand for the average annualised mean, volatility, commodity weights, and portfolio Sharpe ratio of weekly net returns. *, **, *** indicate the significance of a two-tailed significance test at a 10%, 5%, and 1% significance level.

Strategy		Benchmark	GSCI	DBOYBA	SDCI	MSSO	CSMOVERS	CYDMNP	BBLS
EW	Mean	6.86%	6.53%	7.20%	7.54%	6.03%	7.34%	6.96%	7.67%
	StDev	9.74%	9.86%	9.62%	9.55%	8.07%	8.61%	8.70%	8.80%
	Sharpe	0.54	0.5***	0.58***	0.62***	0.55**	0.67***	0.62***	0.69***
5%	Mean		6.84%	7.14%	7.29%	6.61%	7.20%	7.03%	7.35%
	StDev		9.74%	9.67%	9.64%	8.95%	9.20%	9.28%	9.31%
	Sharpe		0.54***	0.57***	0.59***	0.56**	0.61***	0.59***	0.62***
10%	Mean		6.59%	7.19%	7.49%	6.13%	7.32%	6.97%	7.61%
	StDev		9.84%	9.63%	9.58%	8.22%	8.71%	8.81%	8.89%
	Sharpe		0.51***	0.58***	0.62***	0.55**	0.66***	0.61***	0.68***
15%	Mean		6.34%	7.24%	7.70%	5.66%	7.43%	6.91%	7.87%
	StDev		10.05%	9.65%	9.56%	7.60%	8.29%	8.33%	8.51%
	Sharpe		0.47***	0.59***	0.64***	0.54***	0.7***	0.64***	0.74***
Aggr. 15%	Mean	7.56%	6.69%	7.60%	8.05%	6.01%	7.78%	7.26%	8.23%
	StDev	14.31%	13.31%	12.96%	12.88%	10.90%	11.60%	11.73%	11.86%
	Sharpe	0.42	0.38***	0.46***	0.5***	0.41	0.53***	0.48***	0.56***
Cons. 5%	Mean	6.62%	6.33%	6.63%	6.78%	6.10%	6.69%	6.52%	6.84%
	StDev	5.98%	5.78%	5.71%	5.67%	5.08%	5.29%	5.33%	5.38%
	Sharpe	0.84	0.82	0.88**	0.92***	0.89**	0.96***	0.92***	0.98***

The results of the EW and SW strategies for the before launch periods in **Table 15** in the Appendix further corroborate the complete period findings that commodity indices improve the SR of a diversified portfolio.

The results of the EW and SW strategies for the after launch period in **Table 16** in the Appendix however, do not confirm the previous statistically significant findings. After December 2010, the results of all strategies are opposite to the previous ones and now only the MSSO, which previously underperformed the benchmark, is again the only index that significantly improves the benchmark portfolio for all naïve strategies. The SR's for all portfolios are however greater than for the before launch period, which can be mainly attributed to the lower volatility values and the bullish market time. The results confirm the descriptive statistics of the after launch period and further support the evidences that commodities could not improve the return of portfolios but lower the volatility for the after launch period or this bullish phase.

Finally, these results are in line with the previous assumptions of the diversification benefits inherent in the generation commodities, the descriptives and the mv spanning tests. All commodity indices are able to reduce the volatility of a portfolio, independent of the examined period. No indice however is able to increase the portfolio return for the after launch period. Furthermore, only the MSSO is able to improve the SR after the launch period, which confirms its superior findings of the descriptive statistics. Concluding, the diversification benefits for investors is more pronounced due to risk reductions and lesser to return improvements.

5.2.3 IS results of simple and optimisation strategies

In **Table 9** the results of the IS simple and optimisation strategies with annual rebalancing are presented for the complete time period.

In general, most of the IS strategies of the augmented portfolios indicate higher SR's than the benchmark portfolio because of lower volatility values. The GSCI always improves the SR for all strategies, apart from the minVar, in which it does not change it. Surprisingly, most of the time it experiences higher returns than the benchmark but mixed volatility rates, which is contrary to the previous findings. For the 2nd generations, the SR is always significantly higher than of the benchmark and also the 1st generation, which is as expected due lower volatilities or for the maxSR due to higher returns. The 3rd generation indices however show ambivalent results. Only the BBLS always outperforms the benchmark for all strategies. The CYDMNP, for instance, does not do so for the high-risk R2R strategy because of its low risk but also low return profile. The commodity allocation towards CYDMNP and BBLS are on average significantly higher than for the other indices and range from 34% to 63% and for the BBLS from 6% to 25%. Especially the extraordinary low volatility rates of the CYDMNP depicts beneficial diversification features for the low-risk strategies. The CSMOVERS and the MSSO do not outperform the benchmark portfolio for the minVar strategy. For more risky strategies like the maxSR and the R2R higher quantities are allocated to commodities, in general.

Concluding, on average all commodity indices yield positive attributes that very likely improve the SR of a diversified portfolio in-sample. Whereas 1st and 2nd generations show steady and positive outcomes, 3rd generation indices experience arbitrary performances. The CYDMNP and the BBLS perform better for low-risk strategies such as the minVar and RP than the CSMOVERS and the MSSO. The latter ones however perform better for more risky strategies. Moreover, 2nd generations perform better than 1st generations, and only the good performing 3rd generations indices better than 2nd generations.

Table 9: Optimisation and simple IS strategy results over the complete period

Notes: IS results for all applicable strategies over the complete period from 02.2000 until 05.2018. The strategy column refers to the strategy applied, the second column indicates the results, and the Benchmark column depicts the results of the traditional portfolio. Benchmark indicates the results of the traditional portfolio and the indices abbreviations refer to the augmented portfolios including the commodity index. Mean, StDev, Com weight, and Sharpe stand for the average annualised mean, volatility, commodity weights, and Sharpe ratio of weekly net returns. *, **, *** indicate the significance of a two-tailed significance test at a 10%, 5%, and 1% significance level.

Strategy		Benchmark	GSCI	DBOYBA	SDCI	MSSO	CSMOVERS	CYDMNP	BBLS
MinVar	Mean	5.57%	5.48%	5.56%	5.75%	4.87%	5.40%	5.52%	6.10%
	StDev	3.34%	3.27%	3.25%	3.24%	3.13%	3.17%	2.00%	3.10%

	Com weight	-	3.05%	4.86%	5.83%	7.50%	7.59%	62.66%	10.87%
	Sharpe	1.27	1.27	1.32***	1.38***	1.11***	1.25	1.8***	1.48***
MaxSR	Mean	15.71%	17.27%	17.00%	17.48%	16.30%	16.50%	12.08%	17.20%
	StDev	5.56%	6.10%	5.82%	5.77%	6.14%	5.48%	3.40%	5.71%
	Com weight	-	8.36%	13.17%	14.08%	10.21%	11.02%	45.01%	25.43%
	Sharpe	2.51	2.72***	2.76***	2.86***	2.77***	2.7***	3.16***	2.90***
RP	Mean	6.70%	6.82%	6.82%	6.89%	6.49%	6.78%	6.45%	6.95%
	StDev	4.19%	4.19%	4.18%	4.17%	4.02%	4.05%	2.45%	3.98%
	Com weight	-	1.48%	3.39%	3.88%	2.61%	3.52%	49.81%	6.14%
	Sharpe	1.28	1.31***	1.31***	1.33***	1.28***	1.34***	1.82***	1.4***
R2R	Mean	15.22%	15.99%	15.99%	16.33%	17.47%	16.35%	14.09%	16.35%
	StDev	5.57%	5.57%	5.57%	5.57%	5.57%	5.57%	5.57%	5.57%
	Com weight	-	7.30%	10.39%	10.77%	6.95%	6.88%	34.11%	17.64%
	Sharpe	2.30	2.45***	2.45***	2.53***	2.8***	2.53***	2.11***	2.52***

The results of the complete period in **Table 17** in the Appendix confirm the findings of the complete period. **Table 18** in the Appendix however does not confirm the conclusions for the period after the launch. This is in accordance with the previous outcomes, in which the after launch period already signals different performances compared to the other periods. For this period only the minVar SR's show higher values for all portfolios compared to the before launch period. All the other strategies seem to perform worse during the after launch period and show lower SR's, which is in contrast to the static IS strategies that show better performances for the latest period for all strategies. Furthermore, even though the SR's are lower than in the before launch period, surprisingly only the BBLS is able to outperform the benchmark portfolio for all strategies. The MSSO, which usually outperformed the other indices in the static strategy analyses is this time not able to outperform the benchmark for the minVar strategy, due to lower returns. Nonetheless, it shows strong significant results for the other strategies up to a maximum SR of 2.94 for the R2R strategy. The CSMOVERS performs the worst for all strategies and even underperforms 1st and 2nd generations. The CYDMNP shows mixed results and performs best for the maxSR and RP strategies but the worst results of all indices for the minVar and the R2R strategies. The 1st and 2nd generation indices show also inconclusive results and outperform the benchmark rather for the risky strategies but underperform with the low-risk strategies.

In summary, there are again mixed results for the different periods studied. For the complete and before launch periods, each generation commodity index can significantly improve the IS performance of a traditional portfolio either by reducing the volatility or by increasing the return, which depends on the strategy applied. For the period after the launch however, 1st and 2nd generation indices depict homogeneous performances and out-, underperform the benchmark depending on the strategy. Within the 3rd generation indices there exists different behaviours and results for the changing strategies and thus there are inconclusive patterns within that generation. However, the CSMOVERS is the only index,

which underperforms the other indices for all strategies. Nevertheless, 3rd generations almost always lower the volatility of the traditional portfolio.

Finally, the IS analyses indicate valuable insights about the performance of commodities as diversification opportunities and show that every generation commodity indice is able to reduce the risk. The more active managed indices are even able to increase the return. In addition, 2nd generations are superior to 1st and some 3rd generations superior to 2nd generations. The results of the IS analyses however do not yield clear distinctions of the indices or within the generations and hence OOS analyses are examined to generate more realistic results. Next to that, practitioners are more interested if the indices also demonstrate benefits for OOS strategies. Therefore, this thesis investigates the OOS simple and optimisation strategies in the next section.

5.3 OOS results

This section shows the results of the OOS for the minVar, RP, and R2R strategies based on a weekly rolling sample approach with an estimation window of 1 year over the complete, before and after launch periods. Next to that, whereas there are short sell constraints for the RP and the R2R strategies, there a no such restrictions for the minVar. The transaction costs included per rebalancing also cause the returns and SRs to be lower than for the IS analyses.

5.3.1 OOS minVar and simple strategy results

Concluding, the minVar strategy shows negative returns due to the minimising risk strategy and the shorting possibility (Heier Hansen-Tangen & Overaae, 2015). Therefore, for this setting the minVar strategy without short-selling constraints is not an appropriate strategy to analyse the benefits of generation commodity indices. 1st and 2nd generation indices do not add value to a traditional portfolio for OOS analyses and fail to lower the volatility. Only some selected 3rd generation indices like the MSSO and the BBLS are able to outperform the benchmark portfolio for the after-launch period but all 3rd generations in general reduce the portfolio volatility. Lastly, during the last period, which depicts a long bullish market trend, commodity indices are not favourable asset classes for investors, except for the selected outperformers. For the less important and backfilled periods, the complete and the before-launch ones, all 2nd and all, except for and surprisingly only the MSSO, of the 3rd generations indices are able to outperform the benchmark portfolio. In the end, the OOS results show lower performances than the IS results due to higher transaction costs, rebalancing and estimation errors. Yet, it shows a more realistic assessment of the portfolio advantages from commodities (Bessler & Wolff, 2015).

Reasons for the distinctive performances of 3rd generations are due to their core difference, the individually sophisticated strategies, which allow for short positions (Fethke & Prokopczuk, 2018). In contrast to 1st and 2nd generation indices, 3rd generations should in general outperform the former as these only invest in long positions. Yet, 3rd generations are also able to enter short positions and thus can be significantly more exposed to their strategy by shorting commodities that are in contango. Finally, these actively managed 3rd generations are more likely to dedicate strategies that follow the term structure (Fethke & Prokopczuk, 2018). In the end, due to this unique active management feature, there exist huge differences between the performances of 3rd generations within their own generation and with other passive indices.

Table 10 presents the OOS results that are based on a 1-year estimation window with weekly rebalancing and is calculated after transaction costs for the complete, before and after launch periods.
Starting with the GSCI, the 1st generation index does not improve the SR for any strategy during any period but shows even worse performances for the minVar and RP strategies or the same SR for the R2R strategy. Most of the time is does not improve the mean return but increases the volatility. The weights to this index are relatively similar over the strategies and periods. In summary, these findings do not confirm the benefits derived during the IS calculations and the GSCI does not improve the portfolio performance OOS. Daskalaki et al. (2017), Fethke & Prokopczuk (2018), and Kremer (2015) also affirm these outcomes. The GSCI does also not lessen the portfolio volatility.

The results for the 2nd generation indices depict that they significantly perform better than the 1st generation index for the complete period and the before launch period, but worse for the after launch period. Moreover, the SDCI is only able to significantly outperform the benchmark for the complete and before launch period, and the DBOYBA only able to outperform the before launch period benchmark SR. Both indices however statistically significantly underperform the benchmark for the after launch period. The mean return and the volatility values further support these findings. The weights allocated to the 2nd generation indices are substantially higher than for the 1st generation and depending on the strategy it can be easily double as much. Looking at different periods, the weights increase for the period before the last launch, whereas the weights attributed to the index decrease for the after launch period. In general, more weights are allocated to the SDCI, which outperforms the DBOYBA to some extent. Fethke & Prokopczuk (2018) already note that the SDCI is the best performing 2nd generation index in their analysis, which is because of their specific investment strategy that makes use of price momentum and signals of future term structures. Next, all 2nd generation indices fail to reduce the volatility of the portfolios.
In summary, 2nd generation indices show relatively homogeneous results. In the end, even though the results for the complete and before launch periods show advantageous OOS outcomes, the actual after launch calculations show that an investor does not add value to a diversified portfolio by including 2nd generation indices. These findings are surprising and contradict the findings of the IS analyses. Moreover, it opposes the arguments by Daskalaki et al. (2017), Fethke & Prokopczuk (2018), and Kremer (2015), who argue for beneficial investments in 2nd generations. The deviating conclusion in this study might be due to several reasons. Some are for instance that this study explicitly only investigates the real performance of the 2nd generation indices without backfilled data and is updated until May 2018. Furthermore, this study investigates weekly rebalancing, compared to monthly ones and inspects a broader universe of traditional assets.

The analysis concerning the 3rd generation indices exhibits more heterogeneous conclusions. First, the BBLS outperforms the benchmark during the complete and before launch periods, but shows an

outperformance for the R2R strategy at the 10% level and a significant underperformance for the minVar strategy after launch period. Second, the CSMOVERS and the CYDMNP also outperform most of the time during the complete and before launch period but underperform substantially during the after launch period. Finally, the MSSO oppose the other ones and underperforms the benchmark during the complete and before launch periods but statistically significant outperforms the benchmark for the after launch period. Weights allocated to the 3rd generations show that the MSSO and the CSMOVERS show a similar quantity than 2nd generations, whereas the allocation to the CYDMNP and the BBLS present sizeable increases. In particular, the CYDMNP still allocates large amounts to the portfolio ranging from 34% for the R2R to 62% for the minVar after-launch period and greatly exceeds the amount for all the other indices. Besides, all 3rd generations are able to reduce the volatility of the portfolios.
In summary, the OOS results do confirm the results of 3rd generations from the IS and also indicate disparate performances, as can be seen in **Figure 1**. Additionally, it also confirms the conclusion from Fethke & Prokopczuk (2018), who present different performances for 3rd generation indices and divides these into out- and underperformer. Nonetheless, the useful 3rd generation indices in this study are not similar to the ones, which are useful in their study and vice versa. Whereas the MSSO is the only unambiguously outperforming index here and also to some extent the BBLS, the other two indices do not outperform the benchmark during the after launch period. The other researchers however show that the BBLS, the CYDMNP, and the CSMOVERS are the outperformer, whereas the MSSO an underperformer. Reasons for the different conclusions are for instance the shorter period analysed, the usage of the actual performance data only and weekly return data, the more diverse benchmark portfolio, and the bullish market period trend.

Concluding, the minVar strategy shows negative returns due to the minimising risk strategy and the shorting possibility (Heier Hansen-Tangen & Overaae, 2015). Therefore, for this setting the minVar strategy without short-selling constraints is not an appropriate strategy to analyse the benefits of generation commodity indices. 1st and 2nd generation indices do not add value to a traditional portfolio for OOS analyses and fail to lower the volatility. Only some selected 3rd generation indices like the MSSO and the BBLS are able to outperform the benchmark portfolio for the after-launch period but all 3rd generations in general reduce the portfolio volatility. Lastly, during the last period, which depicts a long bullish market trend, commodity indices are not favourable asset classes for investors, except for the selected outperformers. For the less important and backfilled periods, the complete and the before-launch ones, all 2nd and all, except for and surprisingly only the MSSO, of the 3rd generations indices are able to outperform the benchmark portfolio. In the end, the OOS results show lower performances than the IS results due to higher transaction costs, rebalancing and estimation errors. Yet, it shows a more realistic assessment of the portfolio advantages from commodities (Bessler & Wolff, 2015).

Reasons for the distinctive performances of 3rd generations are due to their core difference, the individually sophisticated strategies, which allow for short positions (Fethke & Prokopczuk, 2018). In contrast to 1st and 2nd generation indices, 3rd generations should in general outperform the former as these only invest in long positions. Yet, 3rd generations are also able to enter short positions and thus can be significantly more exposed to their strategy by shorting commodities that are in contango. Finally, these actively managed 3rd generations are more likely to dedicate strategies that follow the term structure (Fethke & Prokopczuk, 2018). In the end, due to this unique active management feature, there exist huge differences between the performances of 3rd generations within their own generation and with other passive indices.

Table 10: OOS results over the complete, before, and after launch periods for all strategies for 1-year window sizes.

Notes: This table presents the weekly rebalancing OOS results for all applicable strategies from a rolling sample approach with a 1-year estimation window over the complete, before, and after launch periods. The period and strategy column refers to the examined period and the employed strategy. The second column indicates the performance measure. Benchmark indicates the results of the traditional portfolio and the indices abbreviations refer to the augmented portfolios including the commodity index. Mean, StDev, Com weight, and SR stand for the average annualised mean, volatility, commodity weights, and Sharpe ratio of the annualised excess returns over the annualised volatility of weekly net returns. *, **, *** indicate the significance of a two-tailed significance test at a 10%, 5%, and 1% significance level.

Period & strategy		Benchmark	GSCI	DBOYBA	SDCI	MSSO	CSMOVERS	CYDMNP	BBLS
Complete									
MinVar	Mean	-1.18%	-1.56%	-1.38%	-1.11%	-1.49%	-1.10%	0.44%	-0.88%
	StDev	4.11%	4.10%	4.08%	4.06%	4.01%	3.87%	2.44%	3.91%
	Com weights		3.15%	5.88%	6.19%	4.65%	5.68%	59.96%	8.42%
	SR	-0.67	-0.77***	-0.73***	-0.67	-0.77***	-0.7*	-0.47***	-0.63***
RP	Mean	4.60%	4.51%	4.62%	4.73%	4.33%	4.65%	3.99%	4.79%
	StDev	4.62%	4.63%	4.64%	4.62%	4.42%	4.44%	2.73%	4.40%
	Com weights		1.47%	3.28%	3.79%	2.46%	3.47%	49.37%	6.36%
	SR	0.65	0.63***	0.65	0.68***	0.62***	0.69***	0.88***	0.73***
R2R	Mean	4.37%	4.34%	4.36%	4.53%	4.37%	4.34%	3.20%	4.73%
	StDev	4.92%	4.93%	4.97%	4.97%	4.92%	4.71%	2.97%	4.55%
	Com weights		0.53%	4.25%	6.43%	0.00%	5.08%	47.87%	11.56%
	SR	0.56	0.56***	0.56	0.59***	0.56	0.58***	0.54	0.69***
Before									
MinVar	Mean	-1.58%	-1.94%	-1.48%	-1.14%	-2.05%	-1.17%	1.78%	-0.54%
	StDev	4.92%	4.92%	4.90%	4.86%	4.80%	4.58%	2.89%	4.64%
	Com weights		4.00%	7.78%	8.52%	4.65%	7.47%	59.26%	10.06%
	SR	-0.82	-0.89***	-0.80	-0.74***	-0.94***	-0.79	-0.23***	-0.65***
RP	Mean	6.17%	6.17%	6.46%	6.59%	5.66%	6.42%	6.20%	6.60%
	StDev	5.16%	5.18%	5.18%	5.16%	4.94%	4.97%	3.13%	4.92%
	Com weights		1.34%	3.48%	3.71%	2.59%	3.69%	44.30%	6.12%
	SR	0.72	0.72	0.77***	0.8***	0.65***	0.8***	1.2***	0.84***
R2R	Mean	6.32%	6.31%	6.78%	7.19%	6.32%	6.72%	5.45%	7.14%
	StDev	5.20%	5.21%	5.29%	5.29%	5.20%	4.97%	3.18%	4.83%
	Com weights		1.23%	6.61%	8.54%	0.00%	7.64%	47.03%	12.44%
	SR	0.74	0.74	0.82***	0.89***	0.74	0.86***	0.94***	0.97***
After									
MinVar	Mean	-1.07%	-1.51%	-1.66%	-1.45%	-0.96%	-1.52%	-1.70%	-1.74%
	StDev	2.54%	2.52%	2.49%	2.52%	2.51%	2.56%	1.56%	2.53%
	Com weights		2.23%	4.09%	3.64%	4.13%	3.02%	61.68%	6.10%
	SR	-0.54	-0.72***	-0.79***	-0.7***	-0.51**	-0.71***	-1.28***	-0.81***
RP	Mean	2.40%	2.17%	2.05%	2.14%	2.45%	2.11%	0.84%	2.25%
	StDev	3.70%	3.70%	3.71%	3.69%	3.57%	3.55%	1.93%	3.52%
	Com weights		1.68%	3.21%	4.25%	2.35%	3.29%	57.04%	6.79%
	SR	0.56	0.5***	0.47***	0.49***	0.6***	0.5***	0.27***	0.55
R2R	Mean	2.74%	2.74%	2.74%	2.74%	2.75%	2.74%	0.94%	2.66%
	StDev	4.80%	4.80%	4.80%	4.80%	4.66%	4.80%	3.35%	4.52%
	Com weights		0.00%	0.00%	0.00%	1.50%	0.00%	34.42%	6.62%
	SR	0.51	0.51	0.51	0.51	0.52***	0.51	0.19***	0.52*

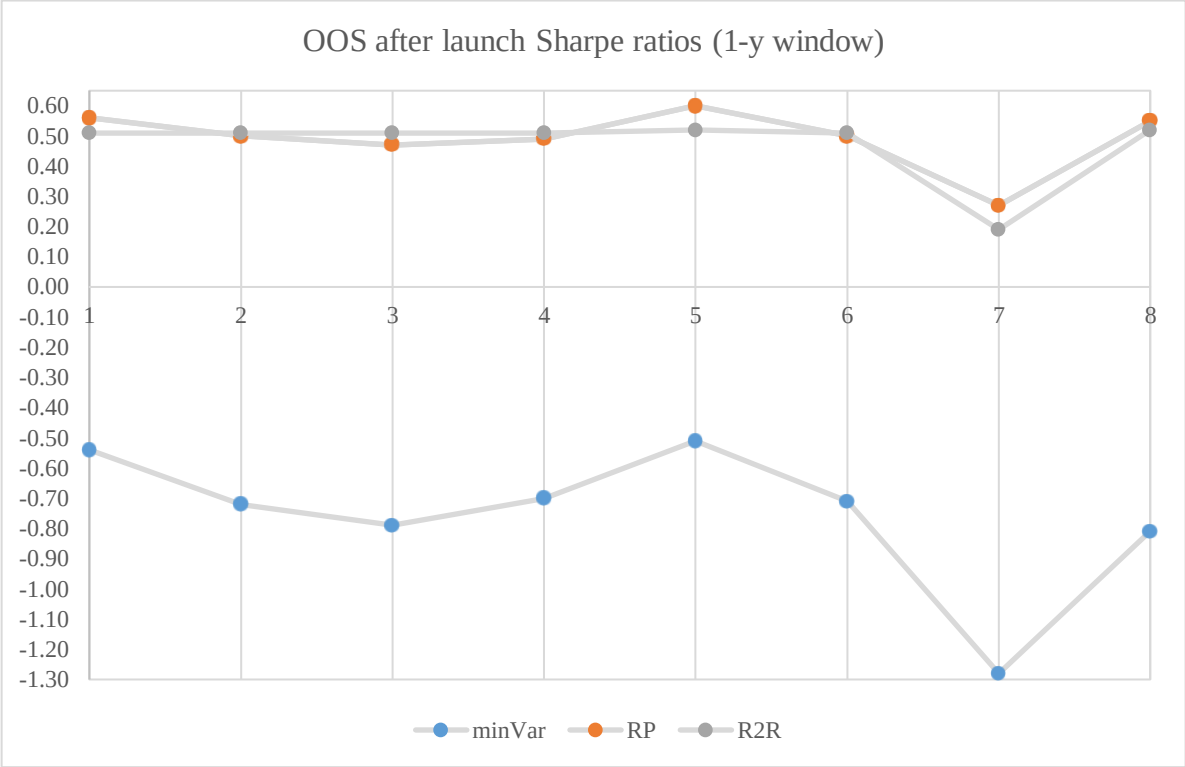


Figure 1: OOS after launch Sharpe ratios (1-y estimation window)

This figure shows the calculated Sharpe ratios for the OOS strategies based on the 1-year window size for the after launch period from December 2010 until May 2018. The y-axis depicts the Sharpe ratio and the x-axis the portfolios, in which 1 depicts the benchmark portfolio, 2 the augment portfolio with the GSCI, 3 the portfolio with the DBOYBA, 4 the SDCI, 5 the MSSO, 6 the CSMOVERS, 7 the CYDMNP and 8 the BBLS augmented portfolio. The graphs show the different strategy performance.

5.3.2 OOS turnover results

In this section, the portfolio turnover is described that shows the average reallocation of assets in the portfolio per weekly rebalancing period.

Table 11 summarises the turnover for all OOS calculations for an annual window size with weekly rebalancing over the complete, before and after launch periods. The results indicate that the turnover increases when commodity indices are included in a portfolio. On average, the outcomes of each strategy show similar values over the different periods. Moreover, the RP strategy displays substantially lower turnover rates than the R2R and the minVar strategy due to the inherent low volatility strategy (Bessler & Wolff, 2015). R2R for instance also accounts for expected asset returns and therefore includes more risky assets, which more likely results in higher reallocation values.

One notable exception in this analysis however is the CYDMNP, which demonstrates significantly lower turnover rates for the minVar strategy. An explanation for this are the large weights allocated to the portfolios and the lower volatility rates compared to the traditional asset classes.

In general, the findings in this study are in line with Fethke & Prokopczuk (2018) and Henriksen (2018), who also show similar results of the turnover ratio for the applied strategies. Furthermore, the higher the turnover rate, the higher the transaction costs and thus the lower the SR. Corresponding, the minVar shows on average lower SR's than the other strategies and the RP higher ones than the R2R during the negative periods due to the inherent importance of lower risks.

Table 11: Portfolio turnover for OOS calculations

Notes: This table shows the average turnover change (in percentage) over all rebalancing points for the weekly rebalancing OOS results with 1-year estimation windows. The strategy column refers to the employed OOS strategy and the second column refers to the analysed period.

Strategy		Benchmark	GSCI	DBOYBA	SDCI	MSSO	CSMOVERS	CYDMNP	BBLS
minVar	Complete	7.80	8.00	8.10	8.13	8.20	7.95	5.91	8.02
	Before	8.82	8.96	9.04	9.08	9.26	8.86	6.56	8.88
	After	6.35	6.61	6.73	6.77	6.69	6.72	5.00	6.76
RP	Complete	1.28	1.29	1.30	1.30	1.29	1.30	1.35	1.34
	Before	1.33	1.34	1.35	1.35	1.34	1.35	1.38	1.37
	After	1.22	1.23	1.24	1.25	1.24	1.24	1.30	1.31
R2R	Complete	2.66	2.67	2.75	2.77	2.66	2.77	2.84	2.79
	Before	2.66	2.67	2.75	2.77	2.66	2.77	2.84	2.79
	After	2.82	2.82	2.82	2.82	2.84	2.82	3.15	2.93

5.4 Robustness checks

This section summarises the results of more robustness tests in order to validate the conclusion of the previous sections and to illustrate more realistic evaluations. In particular, more time periods are researched depending on the different market trends but also different rebalancing periods for the OOS analyses.

5.4.1 Spanning results for different periods

In **Table 19** the test statistics of the GMM-Wald spanning test over two different bearish and two bullish periods are presented.

As can be seen, all 3rd generation indices and the SDCI can reject the null hypothesis of spanning and significantly add value to a portfolio for all periods. Only during the 3rd period the GSCI and the DBOYBA are spanned by the benchmark portfolio and do not add value. The number of observations for this phase however is rather limited and should therefore not be overrated.

The conclusions of these sub-periods endorse the IS findings that generation commodity indices in general are able to reduce the portfolio risks over different market trends. Again, the low correlation to the traditional asset classes allow for promising diversification effects for investors. The OOS conclusion

however suggest that only some 3rd generation indices are able to do so, which is in line with the risk-return profile during the after-launch period of the indices. In the end, Bessler and Wolff (2015) show that IS spanning tests tend to overstate the benefits of commodities investments and thus the spanning test should be assessed with caution and the outcomes of the OOS calculations considered as superior (Fethke & Prokopczuk, 2018).

5.4.2 OOS results with different estimation windows

Table 20 and ***Table 21*** summarise the performances of the OOS strategies for the complete, before, and after launch periods based on 6-months and 2-year window sizes with weekly rebalancing and calculated after transaction costs, respectively. The focus of this section is on the after-launch periods.

In short, the results for the 6-months estimation window in **Table 20** and over the complete period are similar to the 1-year estimation window results. Again, all indices underperform the benchmark portfolio for every strategy. Only the MSSO outperforms the benchmark for the RP and the R2R strategy and the BBLS for the RP at the 10% significance level. Compared to the 1-year estimation window, the minVar and the RP show lower SRs. The R2R strategy however seems to be more suitable here and depicts remarkably better SRs for the last period. All the other indices underperform the benchmark portfolio's SR and there is no difference between 1st and 2nd generations. Additionally, only 3rd generation indices are able to reduce the overall volatility.

The results of the 2-year estimation window in ***Table 21*** however show better performances for the three 3rd generation indices MSSO, CYDMNP and BBLS. As can be seen in **Figure 2**, except for the minVar strategy, for which only the MSSO is able to outperform the benchmark, the other two indices are also able to outperform the benchmark portfolios for the RP and the R2R strategies as well. All the other indices underperform the benchmark portfolio's SR and there is no difference between 1st and 2nd generations. Whereas the SRs of the minVar and the RP are higher compared to the 6-month window, the R2R shows somewhat lower SRs. Compared to the 1-year window values, the SRs are only somewhat lower for the RP strategy. Additionally, all 2nd and 3rd generation indices are able to reduce the overall volatility.

Concluding, after the investigation of different estimation windows, the conclusion of the OOS strategies based on the 1-year estimation windows are substantiated for the other estimation windows and the after-launch period in this study, as well. The MSSO again shows the highest SRs, followed by the BLLS and this time also by the CYDMNP. Only the CSMOVERS underperforms the benchmark again, which is still in contrast to Fethke's & Prokopczuk's conclusion (2018) but in line with the expectations of the descriptive statistics of this paper. 1st and 2nd generations fail to reduce the volatility, but 3rd generations

lower it. Besides, 2-year estimation window sizes can be considered to show better SRs because of the ability to use more observations to calculate optimal strategy weights.

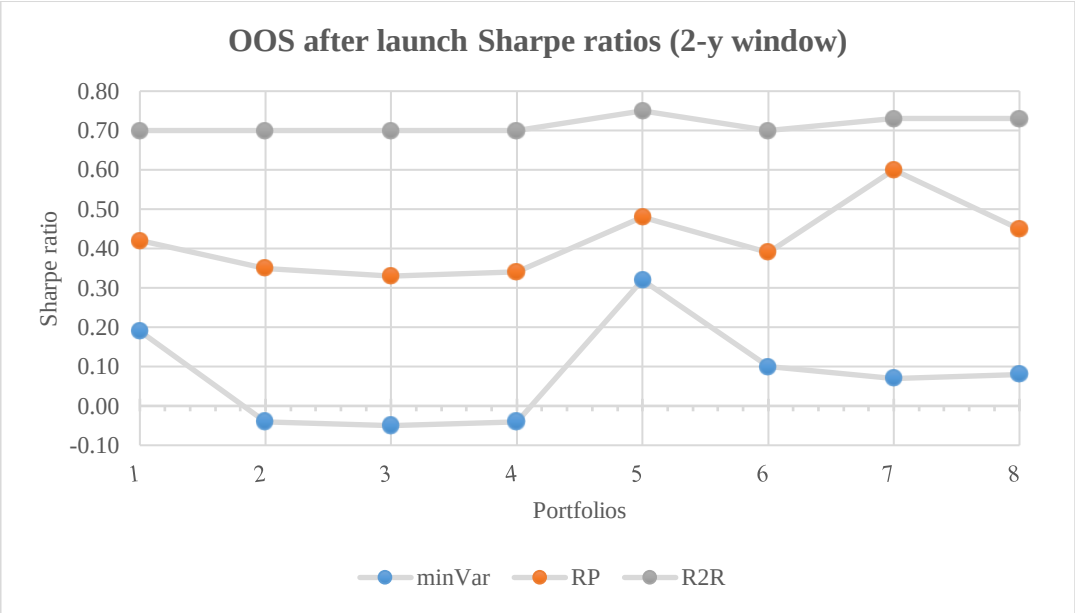


Figure 2: OOS after launch Sharpe ratios (2-y estimation window)

This figure shows the calculated Sharpe ratios for the OOS strategies based on the 2-year window size for the after launch period from December 2010 until May 2018. The y-axis depicts the Sharpe ratio and the x-axis the portfolios, in which 1 depicts the benchmark portfolio, 2 the augment portfolio with the GSCI, 3 the portfolio with the DBOYBA, 4 the SDCI, 5 the MSSO, 6 the CSMOVERS, 7 the CYDMNP and 8 the BBLS augmented portfolio. The graphs show the different strategy performance.

5.5 Limitations and discussion

Due to different (time) constraints this thesis shows some limitations in its analysis, which might be interesting to inspect for future researcher.

Firstly, this study is restricted to only three OOS strategies, but more strategies, such as the Black-Litterman model or the mean-variance strategies with and without short-selling might produce different outcomes. The neglect of varying rebalancing periods such as monthly, (semi-) annual, and biennial instead of only weekly shifts, including different amounts of transaction costs can also show more robust results. Accordingly, different return index data, such as daily, monthly, or annually returns can be interesting to investigate for further research. Deviating estimation window sizes also show to yield different performances and it can be interesting to investigate the most suitable size(s) for each investors preference. Next to that, alternative 3rd generation indices might show different performances than the ones used in this thesis because of the different investment strategies and alternating results of this study. More performance measures might have also been able to show other results, such as the Sortino ratio, which rather accounts for downside risks when there are more positively skewed results instead of using

only the standard deviation for the SR (Rollinger & Hoffman, 2013). In the end however, one of the main issues is the analysis of the different market trends and the actual performances compared to backfilled data performances. Since most of the 3rd generation indices launched only relatively recently, their actual performance has proven to perform differently than for the periods with the backfilled observations. Therefore, the analysis of actual data after the launches is more suitable to generate conclusions about the usefulness of the newest commodity indices. In addition to that, as the analysis of the after-launch period is based on an extraordinary long bullish market time until now, it can be interesting for researchers to examine the actual performance during bearish periods happening in the future, as well.

CHAPTER 6

Summary and conclusion

Since the last two decades, commodity indices became increasingly interesting for asset managers, who are looking for more diversification opportunities after the great financial crisis. Due to low correlation coefficients with traditional asset classes and attractive risk-return profiles, commodity indices suggest being able to provide diversification benefits to conventional portfolios. At the same time, financial engineers have gradually constructed more sophisticated indices with actively managed long-short strategies, next to the deprecated passive long-only indices. Eventually, Miffre (2012), who was one of the first researchers, analyses these new generation indices and differentiates the types in 1st, 2nd, and 3rd generation indices. The generation indices are classified according to their date of launch but differs especially according to the incorporated innovatory investment strategy. In particular, 3rd generation indices comprise long-short strategies, whereas 2nd as well as 1st generation indices rather follow long-only strategies. Eventually, since 1st and 2nd generation indices have empirically shown to add value to a traditional portfolio, only little empirical research has been investigated about the usefulness of 3rd generation indices to diversified portfolios, which is because of their more recent launch and the questionable backfilled data points. Therefore, the main focus of this paper is on the performance of the indices after the last launched indice, the BBLS, after December 2010 in order to exclude backfilled data.

In this thesis seven commodity indices from three different generations and with distinct inherent strategies are employed. The aim of this study is to investigate the usefulness of especially 3rd generation commodity indices to asset managers over different periods based on IS and OOS analyses.

First, this paper conducts IS spanning and step-down tests and shows that all generation indices are able to reject the null of spanning and reduce the volatility for all examined periods. OOS results however fail to produce the results for the 1st and 2nd generation indices.

The results of the naïve IS strategies show that there are differences between the complete period, the before-launch, and the after-launch period. Whereas all, except for the MSSO can outperform the benchmark portfolio for the complete period and only the MSSO is able to outperform the benchmark for the latest period.

The analyses of the four simple and optimisation IS strategies show similar results since not all of the 3rd generation indices are superior to the other two generations again, which exhibit homogeneous patterns. Therefore, there are great performances differences within the 3rd generations as the CSMOVERS shows underperforming results compared to the rest for the after-launch period.

Finally, the OOS analyses depict that 1st and 2nd generation do not add value to a diversified portfolio. 3rd generations again show heterogeneous results and confirm the previous findings. Only the MSSO, the BBLS, and for the last period with 2-year estimation windows also the CYDMNP, are able to significantly add value to a traditional portfolio.

In the end, the answers to the research questions are not clear cut. On the one hand, there are 3rd generation indices, such as the MSSO, the BBLS and sometimes the CYDMNP that can outperform a diversified portfolio and add value to it, depending on the settings. On the other hand, there is a 3rd generation index, the CSMOVERS, which is not able to show similar results and outperforms. In addition to that, the CSMOVERS is not even able to outperform 2nd generation indices, such as the SDCI. Consequently, the previous answers point to similar outcomes of the four hypotheses, which also cannot be unambiguously answered. The hypothesis $H0_a$ for instance can only be rejected for the MSSO, the CYDMNP, and the BBLS but cannot be rejected for the CSMOVERS. Again, the CSMOVERS is not able to produce favourable results to outperform the benchmark portfolio for most of the analyses. Accordingly, the hypothesis $H0_b$ can also only be rejected for the MSSO, the CYDMNP, and the BBLS, which outperform the previous generations because of the same reasons mentioned above. The CSMMOVERS again is not able to reject the null hypothesis. Since the analysis is based on a long bullish market trend the examination of the performance of generation indices during bearish market periods might be interesting for future researchers.

Summarising, the performance of newly invented indices should be evaluated based on actual data points and not on backfilled ones. Moreover, actively managed commodity indices are more likely to outperform passively managed indices and also to add value to a diversified portfolio, this however is not always true per se.

CHAPTER 7
References

Anderson, R. M., Bianchi, S. W., & Goldberg, L. R. (2012). Will my risk parity strategy outperform? *Financial Analysts Journal*, *68*(6), 75–93. http://doi.org/10.2469/faj.v68.n6.7

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2012). Leverage aversion and risk parity. *Financial Analysts Journal*, *68*(1), 47–59. http://doi.org/10.2469/faj.v68.n1.1

Belousova, J., & Dorfleitner, G. (2012). On the diversification benefits of commodities from the perspective of euro investors. *Journal of Banking and Finance*. http://doi.org/10.1016/j.jbankfin.2012.05.003

Benartzi, S., & Thaler, R. H. . (2001). Naive Diversification Strategies in Defined Contribution Saving Plans. *The American Economic Review*, *91*(1), 79–98.

Berndt, B. Y., & Savin, N. (1977). Conflict among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model. *Econometrica*, *45*(5), 1263–1277.

Bessler, W., & Wolff, D. (2015). Do commodities add value in multi-asset portfolios? An out-of-sample analysis for different investment strategies. *Journal of Banking and Finance*, *60*, 1–20. http://doi.org/10.1016/j.jbankfin.2015.06.021

Bjornson, B., & Carter, C. a. (1997). New Evidence on Agricultural Commodity Return Performance under Time-Varying Risk. *American Journal of Agricultural Economics*, *79*(August), 918–930. http://doi.org/10.2307/1244432

Blitz, D., & De Groot, W. (2014). Strategic allocation to commodity factor premiums. *The Journal of Alternative Investments*, *17*(2), 103–116. http://doi.org/10.3905/jai.2014.17.2.103

Bodie, Z., & Rosansky, V. I. (1980). Risk and Return in Commodity Futures. *Financial Analysts Journal*, *36*(3), 27-31+33-39.

Breusch, T. S. (1979). Conclict among criteria for testing hypotheses: Extensions and comments. *Econometrica*, *47*(1), 203–207.

Cheng, I.-H., & Xiong, W. (2014). Financialization of Commodity Markets. *Annual Review of Financial Economics*. http://doi.org/10.1146/annurev-financial-110613-034432

Conover, M., Jensen, G., Johnson, R., & Mercer, J. (2010). *Is Not the Time to Add Commodities to Your Portfolio? Journal of Investing; Fall* (Vol. 19).

Daigler, R. T., Dupoyet, B., & You, L. (2017). Spicing up a portfolio with commodity futures: Still a good recipe? *Journal of Alternative Investments*. http://doi.org/10.3905/jai.2017.19.4.008

Daskalaki, C., & Skiadopoulos, G. (2011). Should investors include commodities in their portfolios after all? New evidence. *Journal of Banking and Finance*, *35*(10), 2606–2626. http://doi.org/10.1016/j.jbankfin.2011.02.022

Daskalaki, C., Skiadopoulos, G., & Topaloglou, N. (2017). Diversification benefits of commodities: A stochastic dominance efficiency approach. *Journal of Empirical Finance*, *44*, 250–269.

http://doi.org/10.1016/j.jempfin.2017.07.004

De Roon, F. A., & Nijman, T. E. (2001). Testing for mean-variance spanning: A survey. *Journal of Empirical Finance*, *8*(2), 111–155. http://doi.org/10.1016/S0927-5398(01)00022-6

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, *22*(5), 1915–1953. http://doi.org/10.1093/rfs/hhm075

Erb, C. B., & Harvey, C. R. (2016). Conquering Misperceptions about Commodity Futures Investing. *Www.Cfapubs.Org Financial Analysts Journal*, *72*(4).

Fethke, T., & Prokopczuk, M. (2018). *Is Commodity Index Investing Profitable?*

Gorton, G., Rouwenhorst, G., & Bhardwaj, G. (2015). Facts and Fantasies about Commodity Futures Ten Years Later. *Yale International Center for Finance*. Retrieved from http://ssrn.com/abstract=2610772

Gorton, G., & Rouwenhorst, K. G. (2006). *Facts and Fantasies about Commodity Futures*. *National Bureau of Economic Research*. Retrieved from http://www.nber.org/papers/w10595

Greer, R. J. (1978). Conservative Commodities. A Key Inflation Hedge. *The Journal of Portfolio Management Summer*, *4*(4), 26–29.

Greer, R. J. (2007). The Role of Commodities in Investment Portfolios. *CFA Institute Conference Proceedings Quarterly*, *24*(4), 35–46.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, *50*(4), 1029–1054.

Heier Hansen-Tangen, N., & Overaae, M. (2015). *Do Commodities Offer Diversification Benefits in Multi-Asset Portfolios ?*

Henriksen, T. E. S. (2018). Properties of Long/Short Commodity Indices in Stock and Bond Portfolios. *The Journal of Alternative Investments*, *20*(4), 51–68.

Huberman, G., & Kandel, S. (1987). Mean-Variance Spanning. *The Journal of Finance*, *42*(4), 873–888.

Jensen, G. R., & Mercer, J. M. (2011). Commodities as an Investment. *The Research Foundation of CFA Institute Literature Review*.

Jobson, J. D., & Korkie, B. M. (1981). Performance Hypothesis Testing with the Sharpe and Treynor Measures. *The Journal of Finance*, *36*(4), 889–908.

Kan, R., & Zhou, G. (2012). Tests of Mean-Variance Spanning. *Annals of Economics and Finance*, *13*(1), 139–187.

Kirby, C., & Ostdiek, B. (2012). It's All in the Timing : Simple Active Portfolio Strategies that Outperform Naive Diversification. *Journal of Financial Quantitative Anal.*, *47*(2), 437–467.

Kremer, P. j. (2015). Comparing three generations of commodity indices: New evidence for portfolio diversification. *Alternative Investment Analyst Review*, 30–43.

Lincke, D.-M. (2016). Commodities as an Asset Class: Delivering Beta and Beyond. *CFA Institute*

Conference Proceedings Quarterly, *33*(1), 1–10.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, *7*(1), 77–91.

Memmel, C. (2003). Performance Hypothesis Testing with the Sharpe Ratio. *Finance Letters*, *1*(1), Available at SSRN: https://ssrn.com/abstract=41258.

Miffre, J. (2012). Comparing First, Second and Third Generation Commodity Indices. *Alternative Investment Analysis Review*, *3*(0), 22–33. http://doi.org/10.2139/ssrn.2132949

Miffre, J. (2015). Long-short commodity investing: A review of the literature. *Journal of Commodity Markets*, *1*(December), 3–13. http://doi.org/10.1016/j.jcomm.2016.01.001

Miffre, J., & Fernandez-Perez, A. (2015). The Case for Long-Short Commodity Investing. *The Journal of Alternative Investments*, *18*(1), 92–105. http://doi.org/10.1063/1.3034259

NBIM. (2015). *The diversification potential of real estate*.

Newey, W. K., & West, K. D. (1987). Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review*, *28*(3), 777–787.

Nijman, T. E., & Swinkels, L. A. P. (2003). *Strategic and Tactical Allocation to Commodities for Retirement Savings Schemes*. *CentER Discussion Paper No. 2003-20*. Retrieved from http://dx.doi.org/10.2139/ssrn.425320

Nijman, T. E., & Swinkels, L. A. P. (2008). Strategic and Tactical Allocation to Commodities for Retirement Savings Schemes. In *The Handbook of Commodity Investing* (pp. 522–547).

Rollinger, T., & Hoffman, S. (2013). Sortino ratio: A better measure of risk. *Futures Magazine*, 40–42.

Skiadopoulos, G. (2012). Investing in commodities: Popular beliefs and misconceptions. *Journal of Asset Management*, *13*, 77–83. http://doi.org/10.1057/jam.2011.35

Xiong, J. X., Ibbotson, R. G., Idzorek, T. M., & Chen, P. (2010). The equal importance of asset allocation and active management. *Financial Analysts Journal*, *66*(2), 1–9.

Yan, L., & Garcia, P. (2017). Portfolio investment: Are commodities useful? *Journal of Commodity Markets*. http://doi.org/10.1016/j.jcomm.2017.10.002

Yau, J. K., Schneeweis, T., Robinson, T. R., & Weiss, L. R. (2007). Alternative Investments Portfolio Management. In *CFA Institute Investment Books* (pp. 477–578).

CHAPTER 8
Appendix A

8.1 Drivers of commodity futures index returns

This part of the thesis analyses the underlying drivers of commodity index returns. It is important for investors that are willing to allocate some wealth in this asset class to become able to correctly forecast the components of total returns in order to not experience losses.

Basically, the total return of a commodity index is the sum of the price (or spot) return and the income return, as in Equation (13)

$$\text{Total Return}_{\text{Com}} = \text{Price return} + (\text{collateral return} + \text{roll return}) \tag{13}$$

The price return represents the percentage change of the future prices that is caused from changes in the underlying spot prices after inflation, so it can be divided into the real price return and the inflation. The income (or carry) return in contrary depicts the collateral return/yield and the roll return/yield. Whereas the former describes the risk-free interest rate earned from investing the full value of the futures contract[9], the latter return is the profit or loss from rolling into a new futures contract through time (Erb & Harvey, 2016; Yau et al., 2007). In the end, it is important for a financial investor to roll over to a new contract in order to avoid the obligations and costs that are linked to the physical or monetary settlements (Greer, 2007).

An influential factor for commodity future pricing is the roll return (Erb & Harvey, 2016), of which the direction and magnitude is dependent on the future spot price and the expected spot price (Miffre, 2012). The fundamentals of the roll return in commodity future pricing have its roots in the underlying concepts of backwardation and contango (Yau et al., 2007).

In short, when a futures market is in backwardation a simple buy-and-hold strategy yields a positive roll return and when it is in contango a negative roll return. If it is in backwardation, for instance, the spot price (or also close future price) is above the expected more distant future price. In this case, the roll return is positive because the future price gradually increases over time to become equal to the spot price. Finally, the index, that is long in the nearby contract and about to expire, needs to roll over to the more distant but lower future price again, so profiting from the long position (Jensen & Mercer, 2011).

[9] If the sum of collateral returns represents 100 percent in the form of T-Bills, then the position is called to be fully collateralised

As can be seen in **Figure 3**, in backwardation the term structure of future contracts shows a downward-sloping curve but an upward-sloping curve in contango. The roll yield is positive in backwardation and negative in contangoed markets.

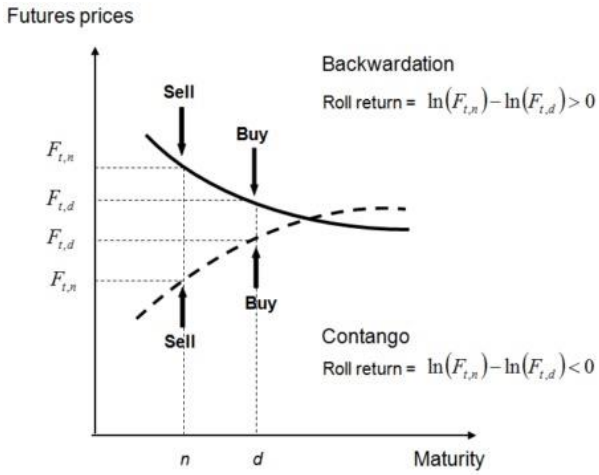


Figure 3: Backwardation and Contango term structure of future contracts (Miffre, 2012)

This figure shows the derivation of the term structure of future contracts. The vertical axis indicates the futures price and the horizontal axis the maturity. The solid line indicates a backwardated roll return and the dashed a contangoed one, in which an investor should sell at time n and buy at time d.

8.2 Details on spanning tests

According to (Belousova & Dorfleitner, 2012; Kan & Zhou, 2012), the LR, Wald, and LM tests can be stated as follows:

$$LR = T \sum_{i=1}^{2} ln(1 + \lambda_i) \overset{A}{\sim} \chi_2^2 \tag{14}$$

$$Wald = T(\lambda_1 + \lambda_2) \overset{A}{\sim} \chi_2^2 \tag{15}$$

$$LM = T \sum_{i=1}^{2} \left(\frac{\lambda_i}{1 + \lambda_i}\right) \overset{A}{\sim} \chi_2^2 \tag{16}$$

All spanning test are regressed against equation (1), assuming normal distribution and estimated using the MLE. The distribution of all three equations is "asymptotically chi-squared with degrees of freedom equal to the number of restrictions under H_0" (Belousova & Dorfleitner, 2012). Again, all three tests need to be performed in order to produce accurate results, since no test is superior to the others (Berndt & Savin, 1977; Breusch, 1979). For a more detailed description of the derivation, see Kan & Zhou (2012).

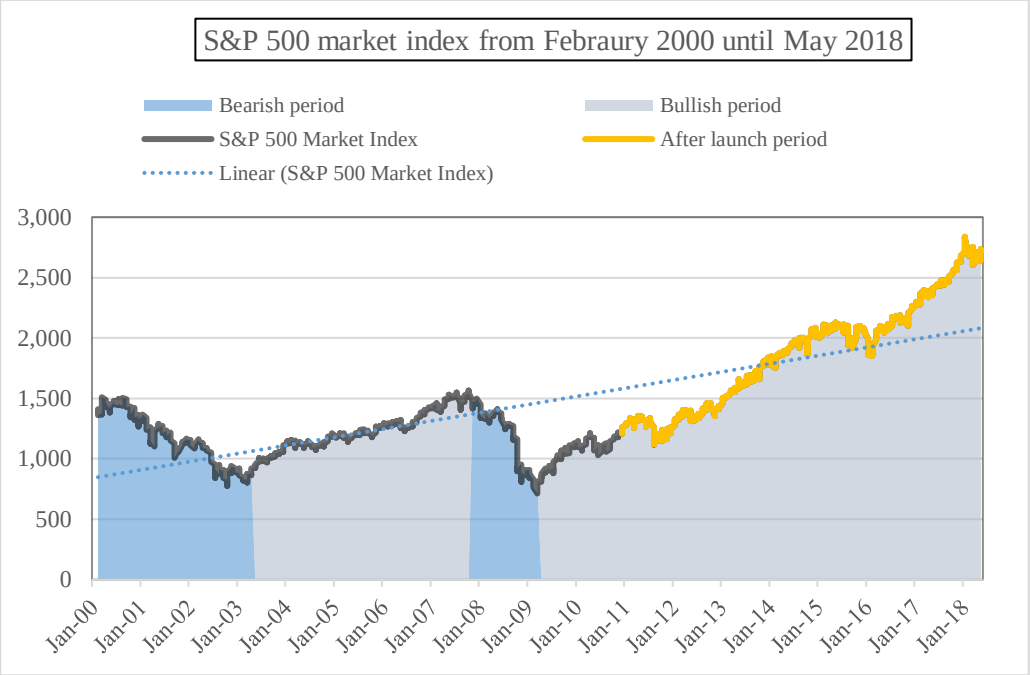
Appendix B - Additional tables

Table 12: Detailed list of commodity indices

Notes*:* This table summarises and describes the commodity indices evaluated in this study. The methodology column describes the strategy, rules, and guidelines. The constituent's column lists the strategy weights and investment position, and also the covered sectors. The base date indicates the pre-inception date, so the hypothetical and back-filled data until the launch date, which refers to the actual date when the index was publicly available, and values equal actual live market data. The information was retrieved from the website of the providers and from Fethke & Prokopczuk (2018), Henriksen (2018), and Miffre (2012).

Datastream/ Bloomberg code	Index name & abbreviations	Strategy	Methodology	Constituent's	Base date	Launch date
1st Gen						
GSCITOT(TR)	S&P Goldman Sachs Commodity Index (GSCI)	Long-only	The relative commodity weights are determined by the world production over the last five years and aim at representing the broad market. The index rolls liquid nearby contracts.	Long-only investment in 24 commodities from all sectors with the emphasis on energy commodities.	31.12.1969	08.01.1991
2nd Gen						
DBLCBBTR	Deutsche Bank Liquid Commodity Index Optimum Yield - Balanced (DBOYBA)	Fundamentals/ Rule-based	The index weights are fixed and determined by commodity world production and inventories of the biggest commodity sectors. It rebalances once a year. Future contracts are selected based on the largest backwardation momentum. The futures rolling dates are rule-based and not predefined, so relying on the implied roll return of eligible contracts that can have maturities up to 13 months. Limited exposure to energy commodities to 35%. The index consists of 14 commodities.	Passive long-only investment in 14 commodities from all sectors, excluding livestock.	03.09.1997	11.01.2007
SDCITR	SummerHaven Dynamic Commodity Index (SDCI)	Fundamentals/ Rule-based	The index is equally weighted of about 7.14% in each index and rebalances on a monthly basis. It takes 7 out of the 27 eligible commodities that have the greatest backwardation. From the remaining 20 commodities, 7 with the greatest 12-month price increase (momentum) are included. The index requires diversification so each sector has to be represented with at least one contract.	Long-only positions in 14 out of 27 eligible commodities from all sectors.	02.01.1991	01.12.2009
3rd Gen						
MSDISTR	Morningstar Short-Only Commodity Index (MSSO)	Momentum	This index engages in fully collaterised short-only momentum strategies. All commodity indices that are ranked in the 95% of the total dollar value of open interest are eligible for short-positions. Cash or long positions are not allowed.	Momentum strategy that engages only in short positions in 20 commodities from all eligible sectors.	21.12.1979	01.08.2007
CSMVRSTR	Credit Suisse Momentum & Volatility Enhanced Return Strategy Index (CSMOVERS)	Momentum	Monthly rebalanced, the index equally invests in 10 out of 24 commodities depending on the return and volatility but must be also included in the SPGSCI. The maturity of the investments depends on the relation of the short-term commodity futures volatility to its long-term average. If the volatility is relatively low, moderate, or high, contracts with six, three, or one month(s) maturity are selected. The price momentum determines if long or short positions are held, depending on positive or negative momentum, respectively. The index exhibits different limits weights for each commodity sector.	Depending on the commodity futures price momentum, the index engages in long or short positions. The index invests in 10 commodity indices from all sectors.	02.01.1998	15.04.2009
CYDIMNTR	CYD Market Neutral Plus Commodity Index (CYDMNP)	Market-Neutral	The index holds position in 15 commodity indices with weights ranging from 3% to 9%. The index is long in the longer end and short in the short end of the term structure of all 15 commodities, given sufficient liquidity. Therefore, it is market-neutral. Returns obtained due to price imbalances during rollover period. Monthly rebalanced.	Market neutral position by holding 15 long and 15 short investment positions in commodities from all sectors.	31.12.1979	01.10.2006
BCCFBKAT	Barclays Backwardation Long Short Index (BBLS)	Term-structure	The index is short in the nearby or front month contracts of the six commodities with the lowest backwardation and at the same time long in the 3-month maturity contracts of the six commodities with the greatest backwardation. The index is based on the 23 commodities represented by Barclays sub-indices.	Six long and six short investment positions in commodities from all sectors.	08.01.1999	26.11.2010

Figure 4: S&P 500 Market Index



This figure shows the S&P 500 market index from February 2000 until May 2018. Furthermore, it marks the bearish and bullish periods in light blue or light grey colours. The yellow line indicates the index for the period after the launch of the last commodity index in December 2010. The blurred line indicates the market trend.

Table 13 Spanning test results of the period before the last launch

Notes: This table presents the test statistics and p-values (in parentheses) of the spanning and step-down test results for all commodity indices over the period before the last launch. The index column specifies the commodity index that is examined against the benchmark portfolio consisting of the 4 bonds and stocks. The columns Wald, LR, and LM and GMM-Wald refer to the Wald, Likelihood ratio, Lagrange multiplier and GMM-based Wald test. F_1 and F_2 present the step-down procedure and indicate if the source of performance enhancement is either due to return improvements or risk reductions, respectively. The examination is based on weekly excess returns over the 3-month US T-Bill risk-free rate. The observations cover the appropriate period as indicated in

Table *3*.

Index	Wald	LR	LM	GMM-Wald	F1	F2
GSCI	21.99	21.57	21.16	16.99	0.11	21.56
	(0.00)	(0.00)	(0.00)	(0.00)	(0.74)	(0.00)
DBOYBA	30.75	29.94	29.16	19.40	4.37	25.73
	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)	(0.00)
SDCI	46.57	44.75	43.02	29.38	10.28	34.96
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
MSSO	38.47	37.21	36.01	29.77	1.67	36.14
	(0.00)	(0.00)	(0.00)	(0.00)	(0.20)	(0.00)
CSMOVERS	70.07	66.05	62.33	49.53	10.23	57.77

	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
CYDMNP	750.74	477.33	322.05	966.04	34.81	663.59
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BBLS	74.31	69.80	65.66	77.46	15.68	55.97
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 14 Spanning test results over the period after the last launch

Notes: This table presents the test statistics and p-values (in parentheses) of the spanning and step-down test results for all commodity indices over the period after the last launch. The index column specifies the commodity index that is examined against the benchmark portfolio consisting of the 4 bonds and stocks. The columns Wald, LR, and LM and GMM-Wald refer to the Wald, Likelihood ratio, Lagrange multiplier and GMM-based Wald test. F_1 and F_2 present the step-down procedure and indicate if the source of performance enhancement is either due to return improvements or risk reductions, respectively. The examination is based on weekly excess returns over the 3-month US T-Bill risk-free rate. The observations cover the appropriate period as indicated in

Table *3*.

Index	Wald	LR	LM	GMM-Wald	F1	F2
GSCI	18.37	17.95	17.55	17.49	2.81	15.07
	(0.00)	(0.00)	(0.00)	(0.00)	(0.09)	(0.00)
DBOYBA	20.79	20.25	19.74	18.67	2.31	17.94
	(0.00)	(0.00)	(0.00)	(0.00)	(0.13)	(0.00)
SDCI	16.24	15.92	15.60	13.06	1.67	14.17
	(0.00)	(0.00)	(0.00)	(0.00)	(0.20)	(0.00)
MSSO	16.53	16.19	15.86	16.10	1.53	14.60
	(0.00)	(0.00)	(0.00)	(0.00)	(0.22)	(0.00)
CSMOVERS	18.54	18.12	17.70	16.04	0.08	18.08
	(0.00)	(0.00)	(0.00)	(0.00)	(0.78)	(0.00)
CYDMNP	636.65	378.18	242.62	772.88	2.54	617.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.11)	(0.00)
BBLS	16.99	16.63	16.29	17.23	0.60	16.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.44)	(0.00)

Table 15 Naïve diversification results over the before launch period

Notes: This table reports the naïve diversification strategies results over the period before the launch of the last commodity index from February 2000 until November 2010. In particular, it reports the EW and SW strategy results. Benchmark refers to the diversified portfolio without commodities. 5%, 10%, 15% refer to the percentage contribution towards commodities. *Aggr*. refers to a portfolio with an aggressive SW investor, *Cons*. refers to a conservative SW investor. Benchmark indicates the results of the traditional portfolio and the indices abbreviations refer to the augmented portfolios including the commodity index. Mean, StDev, and Sharpe stand for the average annualised mean, volatility, and Sharpe ratio of weekly net returns. *, **, *** indicate the significance of a two-tailed significance test at a 10%, 5%, and 1% significance level.

Strategy		Benchmark	GSCI	DBOYBA	SDCI	MSSO	CSMOVERS	CYDMNP	BBLS
EW	Mean	7.61%	7.56%	8.51%	8.99%	6.11%	8.77%	7.78%	8.82%
	StDev	10.83%	10.87%	10.64%	10.60%	9.04%	9.61%	9.66%	9.80%

	Sharpe	0.48	0.47	0.57***	0.62***	0.4***	0.66***	0.55***	0.65***
5%	Mean		7.59%	8.02%	8.23%	6.93%	8.13%	7.69%	8.16%
	StDev		10.77%	10.71%	10.69%	9.97%	10.24%	10.30%	10.34%
	Sharpe		0.48	0.52***	0.54***	0.45***	0.55***	0.51***	0.55***
10%	Mean		7.57%	8.42%	8.85%	6.25%	8.65%	7.76%	8.70%
	StDev		10.84%	10.64%	10.61%	9.19%	9.71%	9.77%	9.89%
	Sharpe		0.47	0.56***	0.6***	0.41***	0.64***	0.54***	0.63***
15%	Mean		7.55%	8.83%	9.48%	5.58%	9.18%	7.83%	9.25%
	StDev		11.03%	10.64%	10.58%	8.53%	9.28%	9.25%	9.48%
	Sharpe		0.46	0.6***	0.66***	0.37***	0.72***	0.58***	0.72***
Aggr. 15%	Mean	7.25%	7.28%	8.55%	9.20%	5.32%	8.91%	7.56%	8.98%
	StDev	15.93%	14.66%	14.34%	14.30%	12.26%	12.99%	13.06%	13.23%
	Sharpe	0.3	0.33**	0.43***	0.47***	0.23***	0.5***	0.39***	0.49***
Cons. 5%	Mean	7.97%	7.98%	8.41%	8.62%	7.32%	8.52%	8.08%	8.55%
	StDev	6.58%	6.35%	6.27%	6.24%	5.59%	5.84%	5.86%	5.93%
	Sharpe	0.84	0.87	0.95***	0.99***	0.87	1.04***	0.96***	1.03***

Table 16 Naïve diversification IS results over the after launch period

Notes: This table reports the naïve diversification strategies results over the period after the launch of the last commodity index from December 2010 until May 2018. In particular, it reports the equally-weighted (EW) and strategically-weighted (SW) strategy results. Benchmark refers to the diversified portfolio without commodities. 5%, 10%, 15% refer to the percentage contribution towards commodities. *Aggr*. refers to a portfolio with an aggressive SW investor, *Cons*. refers to a conservative SW investor. Benchmark indicates the results of the traditional portfolio and the indices abbreviations refer to the augmented portfolios including the commodity index. Mean, StDev, and Sharpe stand for the average annualised mean, volatility, and Sharpe ratio. *, **, *** indicate the significance of a two-tailed significance test at a 10%, 5%, and 1% significance level.

Strategy		Benchmark	GSCI	DBOYBA	SDCI	MSSO	CSMOVERS	CYDMNP	BBLS
EW	Mean	7.25%	5.75%	6.09%	6.22%	6.78%	6.17%	6.59%	6.80%
	StDev	8.00%	8.29%	8.01%	7.89%	6.47%	6.91%	7.14%	7.15%
	Sharpe	0.87	0.66***	0.72***	0.75***	1***	0.85	0.88***	0.91***
5%	Mean		6.57%	6.73%	6.79%	7.04%	6.77%	6.95%	7.05%
	StDev		8.08%	7.98%	7.93%	7.27%	7.47%	7.61%	7.60%
	Sharpe		0.77***	0.8***	0.82***	0.93***	0.86	0.87***	0.89***
10%	Mean		5.90%	6.21%	6.32%	6.83%	6.28%	6.65%	6.85%
	StDev		8.24%	8.00%	7.89%	6.61%	7.00%	7.22%	7.23%
	Sharpe		0.68***	0.74***	0.76***	0.99***	0.85	0.88***	0.9***
15%	Mean		5.22%	5.69%	5.86%	6.62%	5.80%	6.35%	6.64%
	StDev		8.47%	8.05%	7.89%	6.05%	6.60%	6.83%	6.89%
	Sharpe		0.58***	0.67***	0.7***	1.04***	0.83**	0.88***	0.92***
Aggr. 15%	Mean	9.14%	6.61%	7.08%	7.25%	8.02%	7.19%	7.75%	8.04%
	StDev	11.63%	11.14%	10.72%	10.57%	8.63%	9.24%	9.55%	9.57%
	Sharpe	0.76	0.56***	0.63***	0.66***	0.89***	0.74	0.78**	0.81***
Cons. 5%	Mean	5.40%	4.58%	4.73%	4.79%	5.04%	4.77%	4.95%	5.05%
	StDev	4.99%	4.85%	4.79%	4.73%	4.22%	4.37%	4.46%	4.46%
	Sharpe	1.02	0.88***	0.92***	0.95**	1.12***	1.02	1.04	1.06

Table 17: Optimisation and simple IS strategy results over the before launch period

Notes: IS results for all applicable strategies over the before launch period. The strategy column refers to the strategy applied, the second column indicates the performance measure. Benchmark indicates the results of the traditional portfolio and the indices abbreviations refer to the augmented portfolios including the commodity index. Mean, StDev, Com weight, and Sharpe stand for the average annualised mean, volatility, commodity weights, and Sharpe ratio of weekly net returns. *, **, *** indicate the significance of a two-tailed significance test at a 10%, 5%, and 1% significance level.

Strategy		Benchmark	GSCI	DBOYBA	SDCI	MSSO	CSMOVERS	CYDMNP	BBLS
MinVar	Mean	6.88%	6.78%	7.02%	7.31%	6.35%	7.05%	8.37%	7.93%
	StDev	3.76%	3.66%	3.63%	3.62%	3.60%	3.59%	2.46%	3.52%

	Com weight	-	3.32%	6.28%	7.03%	5.64%	7.89%	56.33%	10.87%
	Sharpe	1.09	1.14***	1.25***	1.34***	1***	1.22***	2.3***	1.45***
MaxSR	Mean	20.12%	21.05%	21.34%	21.94%	21.07%	21.81%	14.93%	21.58%
	StDev	6.45%	6.43%	6.34%	6.31%	6.56%	6.55%	3.39%	5.84%
	Com weight	-	5.34%	12.11%	15.23%	9.81%	16.81%	56.78%	27.02%
	Sharpe	2.64	2.85***	2.96***	3.1***	2.88***	2.95***	3.54***	3.16***
RP	Mean	8.84%	9.13%	9.13%	9.29%	8.45%	9.17%	9.12%	9.33%
	StDev	4.55%	4.54%	4.53%	4.51%	4.37%	4.41%	2.90%	4.34%
	Com weight	-	1.38%	3.54%	3.76%	2.73%	3.79%	43.74%	6.06%
	Sharpe	1.43	1.51***	1.52***	1.57***	1.38***	1.55***	2.23***	1.61***
R2R	Mean	19.04%	19.53%	20.43%	21.88%	16.12%	18.53%	16.08%	19.79%
	StDev	5.92%	5.96%	5.94%	5.94%	5.92%	5.89%	5.74%	5.86%
	Com weight	-	3.26%	7.75%	9.11%	5.11%	10.36%	39.28%	14.81%
	Sharpe	2.15	2.37***	2.37***	2.51***	2.53***	2.53***	2.09**	2.48***

Table 18: Optimisation and simple IS strategy results over the after launch period

Notes: IS results for all applicable strategies over the after launch period. The strategy column refers to the strategy applied, the second column indicates the performance measure. Benchmark indicates the results of the traditional portfolio and the indices abbreviations refer to the augmented portfolios including the commodity index. Mean, StDev, Com weight, and Sharpe stand for the average annualised mean, volatility, commodity weights, and Sharpe ratio of weekly net returns. *, **, *** indicate the significance of a two-tailed significance test at a 10%, 5%, and 1% significance level.

Strategy		**Benchmark**	**GSCI**	**DBOYBA**	**SDCI**	**MSSO**	**CSMOVERS**	**CYDMNP**	**BBLS**
MinVar	Mean	4.11%	4.04%	3.93%	4.02%	3.23%	3.58%	2.36%	4.07%
	StDev	2.86%	2.82%	2.83%	2.81%	2.60%	2.71%	1.49%	2.64%
	Com weight	-	2.74%	3.28%	4.49%	9.56%	7.24%	69.69%	10.87%
	Sharpe	1.46	1.42***	1.39***	1.42***	1.23***	1.29***	1.24***	1.51*
MaxSR	Mean	9.65%	12.08%	11.04%	11.35%	9.73%	9.18%	8.15%	11.18%
	StDev	4.33%	5.63%	5.09%	5.04%	5.57%	4.01%	3.43%	5.53%
	Com weight	-	12.50%	13.00%	12.50%	10.77%	3.05%	28.84%	23.24%
	Sharpe	2.33	2.53***	2.50***	2.54***	2.63***	2.35	2.64***	2.54***
RP	Mean	3.77%	3.64%	3.64%	3.60%	3.79%	3.50%	2.77%	3.68%
	StDev	3.69%	3.70%	3.71%	3.69%	3.53%	3.56%	1.84%	3.48%
	Com weight	-	1.61%	3.18%	4.05%	2.44%	3.16%	58.15%	6.25%
	Sharpe	1.08	1.03***	1.02***	1.01***	1.13***	1.04***	1.25***	1.12***
R2R	Mean	11.26%	11.42%	11.42%	11.18%	13.84%	11.10%	9.73%	11.27%
	StDev	5.13%	5.13%	5.13%	5.13%	5.13%	5.13%	5.13%	5.13%
	Com weight	-	0.85%	1.86%	0.76%	9.71%	1.65%	13.86%	9.40%
	Sharpe	2.27	2.29***	2.29***	2.25***	2.94***	2.25***	1.92***	2.31**

Table 19: Spanning test results over 4 alternating market trend periods

Notes: This table presents the test statistics and p-values of the GMM-Wald spanning test results for all commodity indices over the different trend market periods and differentiated in the period column. Period 1 and 3 present bearish and period 2 and 4 bullish periods, as in

Table *3*. The index columns specify the commodity index that is examined against the benchmark portfolio consisting of the 4 bonds and stocks. For all periods, the GMM-based Wald test is shown. The p-values are shown in parentheses. The examination is based on weekly excess returns over the 3-month US T-Bill risk-free rate.

Period	**GSCI**	**DBOYBA**	**SDCI**	**MSSO**	**CSMOVERS**	**CYDMNP**	**BBLS**
1st period	11.49	21.50	24.04	10.24	14.83	179.79	17.44
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
2nd period	5.34	12.07	13.78	11.22	17.79	372.33	23.19
	(0.08)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
3rd period	4.57	2.46	5.86	5.50	14.36	331.76	21.23
	(0.13)	(0.32)	(0.08)	(0.08)	(0.00)	(0.00)	(0.00)
4th period	26.92	23.93	20.26	23.62	27.49	885.59	28.74
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 20: Optimisation and simple OOS strategy results over the complete, before and after launch periods for 6-month estimation windows.

Notes: This table presents the weekly rebalancing OOS results for all applicable strategies from a rolling sample approach with a 6-months estimation window over the complete, before, and after launch periods. The period and strategy column refers to the examined period and the employed strategy. The second column indicates the performance measure. Benchmark indicates the results of the traditional portfolio and the indices abbreviations refer to the augmented portfolios including the commodity index. Mean, StDev, Com weight, and SR stand for the average annualised mean, volatility, commodity weights, and Sharpe ratio of the annualised excess returns over the annualised volatility of weekly net returns. *, **, *** indicate the significance of a two-tailed significance test at a 10%, 5%, and 1% significance level.

Period & strategy		Benchmark	GSCI	DBOYBA	SDCI	MSSO	CSMOVERS	CYDMNP	BBLS
Complete									
MinVar	Mean	-8.88%	-9.89%	-9.76%	-9.45%	-9.88%	-9.42%	-6.28%	-8.96%
	StDev	4.65%	4.72%	4.67%	4.66%	4.63%	4.66%	3.09%	4.62%
	Com weights		3.13%	5.70%	6.07%	4.21%	5.40%	59.50%	7.82%
	SR	-2.25	-2.43***	-2.43***	-2.37***	-2.48***	-2.36***	-2.54***	-2.28
RP	Mean	3.98%	3.90%	4.00%	4.10%	3.66%	3.97%	3.01%	4.20%
	StDev	4.53%	4.54%	4.55%	4.53%	4.35%	4.40%	2.77%	4.31%
	Com weights		1.45%	3.29%	3.75%	2.45%	3.46%	48.89%	6.34%
	SR	0.53	0.51***	0.53	0.55***	0.47***	0.54***	0.51	0.6***
R2R	Mean	5.07%	5.03%	5.07%	5.26%	5.07%	5.12%	4.20%	5.35%
	StDev	5.02%	5.03%	5.07%	5.07%	5.02%	4.75%	2.91%	4.64%
	Com weights		0.54%	4.26%	6.56%	0.00%	5.13%	48.57%	11.80%
	SR	0.69	0.69***	0.69*	0.72***	0.69	0.74***	0.9***	0.81***
Before									
MinVar	Mean	-10.18%	-11.37%	-11.11%	-10.65%	-11.66%	-10.58%	-6.28%	-9.52%
	StDev	5.50%	5.58%	5.48%	5.48%	5.45%	5.52%	3.63%	5.46%
	Com weights		3.98%	7.66%	8.47%	3.94%	6.73%	58.07%	8.49%
	SR	-2.30	-2.48***	-2.47***	-2.39***	-2.59***	-2.36**	-2.41**	-2.19***
RP	Mean	5.71%	5.72%	5.98%	6.11%	5.15%	5.90%	5.06%	6.12%
	StDev	5.00%	5.01%	5.01%	4.99%	4.80%	4.86%	3.16%	4.77%
	Com weights		1.31%	3.50%	3.66%	2.56%	3.57%	43.66%	5.79%
	SR	0.65	0.65	0.7***	0.73***	0.56***	0.71***	0.82***	0.77***
R2R	Mean	6.76%	6.73%	7.23%	7.66%	6.76%	7.28%	6.60%	7.62%
	StDev	5.41%	5.42%	5.50%	5.51%	5.41%	5.05%	3.13%	5.00%
	Com weights		1.26%	6.66%	8.75%	0.00%	7.95%	47.94%	13.29%
	SR	0.80	0.79**	0.87***	0.95***	0.80	0.96***	1.32***	1.03***
After									
MinVar	Mean	-7.92%	-8.97%	-8.91%	-8.63%	-8.37%	-8.94%	-6.92%	-9.08%
	StDev	3.10%	3.13%	3.18%	3.16%	3.14%	3.10%	2.11%	3.08%
	Com weights		2.22%	3.62%	3.29%	4.07%	3.61%	61.93%	6.96%
	SR	-2.66	-2.96***	-2.91***	-2.84***	-2.76***	-2.98***	-3.44***	-3.05***
RP	Mean	1.38%	1.15%	1.01%	1.06%	1.44%	1.03%	-0.12%	1.39%
	StDev	3.83%	3.83%	3.85%	3.83%	3.67%	3.66%	2.07%	3.61%
	Com weights		1.66%	3.12%	4.03%	2.34%	3.36%	56.56%	7.28%
	SR	0.28	0.22***	0.18***	0.19***	0.31***	0.19***	-0.21***	0.3*
R2R	Mean	4.18%	4.18%	4.18%	4.18%	4.18%	4.18%	2.42%	3.92%
	StDev	4.49%	4.49%	4.49%	4.49%	4.37%	4.49%	3.03%	4.25%
	Com weights		0.00%	0.00%	0.00%	1.53%	0.00%	34.07%	6.23%
	SR	0.86	0.86	0.86	0.86	0.89***	0.86	0.7***	0.85

Table 21: Optimisation and simple OOS strategy results over the complete, before and after launch periods for 2-year estimation windows.

Notes: This table presents the weekly rebalancing OOS results for all applicable strategies from a rolling sample approach with a 2-year estimation window over the complete, before, and after launch periods. The period and strategy column refers to the examined period and the employed strategy. The second column indicates the performance measure. Benchmark indicates the results of the traditional portfolio and the indices abbreviations refer to the augmented portfolios including the commodity index. Mean, StDev, Com weight, and SR stand for the average annualised mean, volatility, commodity weights, and Sharpe ratio of the annualised excess returns over the annualised volatility of weekly net returns. *, **, *** indicate the significance of a two-tailed significance test at a 10%, 5%, and 1% significance level.

Period & strategy		Benchmark	GSCI	DBOYBA	SDCI	MSSO	CSMOVERS	CYDMNP	BBLS
Complete									
MinVar	Mean	2.14%	1.97%	2.30%	2.42%	1.77%	2.37%	2.46%	2.50%
	StDev	3.74%	3.72%	3.69%	3.67%	3.66%	3.57%	2.31%	3.59%
	Com weights		3.18%	5.88%	6.21%	4.92%	6.02%	59.89%	8.71%
	SR	0.15	0.1***	0.19***	0.23***	0.05***	0.22***	0.37***	0.25***
RP	Mean	5.52%	5.45%	5.65%	5.70%	5.19%	5.53%	4.56%	5.73%
	StDev	4.60%	4.62%	4.63%	4.61%	4.39%	4.42%	2.70%	4.36%
	Com weights		1.46%	3.21%	3.72%	2.46%	3.46%	49.80%	6.48%
	SR	0.85	0.84***	0.88***	0.89***	0.82***	0.89***	1.1***	0.95***
R2R	Mean	6.06%	6.03%	6.19%	6.32%	6.06%	6.03%	4.79%	6.37%
	StDev	5.01%	5.02%	5.07%	5.08%	5.01%	4.73%	2.89%	4.60%
	Com weights		0.54%	4.20%	6.47%	0.00%	5.16%	49.17%	12.17%
	SR	0.89	0.88***	0.91***	0.93***	0.89	0.94***	1.11***	1.04***
Before									
MinVar	Mean	2.64%	2.69%	3.28%	3.43%	1.93%	3.31%	3.92%	3.55%
	StDev	4.48%	4.46%	4.42%	4.38%	4.37%	4.22%	2.74%	4.27%
	Com weights		3.79%	7.52%	8.04%	5.16%	7.95%	59.22%	10.50%
	SR	0.04	0.05	0.19***	0.22***	-0.12***	0.2***	0.53***	0.26***
RP	Mean	7.47%	7.52%	7.93%	7.99%	6.80%	7.67%	6.81%	7.90%
	StDev	5.21%	5.23%	5.24%	5.22%	4.97%	5.00%	3.17%	4.96%
	Com weights		1.37%	3.52%	3.70%	2.67%	3.84%	44.51%	6.46%
	SR	0.96	0.97	1.04***	1.06***	0.88***	1.04***	1.37***	1.1***
R2R	Mean	8.07%	8.11%	8.92%	9.27%	8.07%	8.42%	7.09%	8.94%
	StDev	5.47%	5.50%	5.60%	5.61%	5.47%	5.10%	3.19%	5.07%
	Com weights		1.30%	6.76%	8.80%	0.00%	8.32%	48.32%	14.13%
	SR	1.03	1.03	1.15***	1.21***	1.03	1.17***	1.45***	1.28***
After									
MinVar	Mean	0.81%	0.22%	0.19%	0.23%	1.14%	0.57%	0.43%	0.53%
	StDev	2.60%	2.52%	2.51%	2.54%	2.60%	2.57%	1.58%	2.55%
	Com weights		2.68%	4.85%	4.66%	3.98%	3.52%	61.10%	5.93%
	SR	0.19	-0.04***	-0.05***	-0.04***	0.32***	0.1***	0.07**	0.08***
RP	Mean	1.90%	1.63%	1.55%	1.60%	2.08%	1.75%	1.46%	1.91%
	StDev	3.81%	3.79%	3.80%	3.79%	3.67%	3.65%	1.89%	3.57%
	Com weights		1.62%	3.22%	4.38%	2.31%	3.25%	56.94%	6.44%
	SR	0.42	0.35***	0.33***	0.34***	0.48***	0.39***	0.6***	0.45***
R2R	Mean	3.50%	3.50%	3.50%	3.50%	3.61%	3.50%	2.56%	3.43%
	StDev	4.55%	4.55%	4.55%	4.55%	4.41%	4.55%	3.07%	4.25%
	Com weights		0.00%	0.00%	0.00%	1.53%	0.00%	33.80%	5.96%
	SR	0.70	0.70	0.70	0.70	0.75***	0.70	0.73*	0.73***