Hannover Center of Finance e.V.

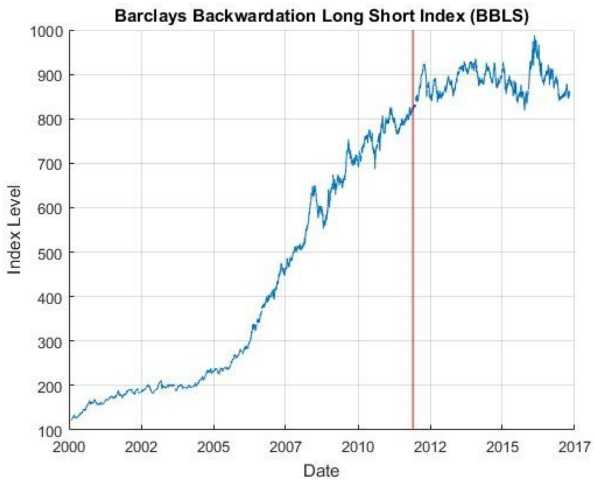
Königsworther Platz 1, 30167 Hannover

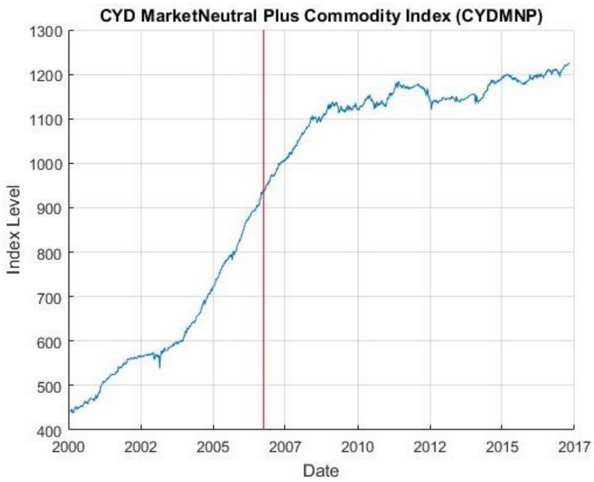
Diskussionspapiere / Discussion Papers
6 (5. April, 2018)[1]

ISSN 2364-2858

Is Commodity Index Investing Profitable?

Tobias Fethke und Marcel Prokopczuk[2]





[1] Kopien oder ein PDF sind erhältlich vom Hannover Center of Finance e.V., Königsworther Platz 1, 30167 Hannover, www.hcf.uni-hannover.de

[2] Institut für Finanzmarkttheorie, Leibniz Universität Hannover, Königsworther Platz 1, 30167 Hannover. Kontakt: prokopczuk@fmt.uni-hannover.de

Is Commodity Index Investing Profitable?*

Tobias Fethke† and Marcel Prokopczuk†,‡

Abstract

Using a comprehensive dataset of first, second and third generation commodity indices, we investigate the potential diversification benefits of commodities in equity–bond portfolios. To this end, different approaches of mean–variance spanning tests and out-of-sample portfolio optimization are implemented. The results show that first generation commodity indices fail to diversify traditional portfolios and are outperformed by enhanced indices. The diversification benefits of second generation indices are indicated by slightly increased portfolio Sharpe ratios but at the same time challenged as they are spanned by benchmark assets. For third generation commodity indices, the mean–variance spanning hypothesis is rejected but they show heterogenous out-of-sample performances. We thus present new evidence showing that the portfolio performance of the third generation of commodity indices is less clear-cut than found in existing studies.

JEL classification: G11

Keywords: Commodities, Investing, Index

*Contact: tobias.fe@web.de (T. Fethke), prokopczuk@fmt.uni-hannover.de (M. Prokopczuk).

†School of Economics and Management, Leibniz University Hannover, Koenigsworther Platz 1, 30167 Hannover, Germany.

‡ICMA Centre, Henley Business School, University of Reading, Reading, RG6 6BA, UK.

1 Introduction

Commodity markets have existed for centuries, representing means for merchants to trade their products and hedge their businesses. Only recently have these markets also moved into the spotlight of private and institutional investors, as indicated by estimates of total investments in commodity index products, which grew from roughly $ 50 bn in 2004 to over $ 300 bn in 2010 (Irwin & Sanders, 2011). A part of this development leads back to evidence presented in the literature that qualifies commodity futures as highly attractive portfolio diversifiers. In fact, they are said to have equity like returns (Gorton & Rouwenhorst, 2006), low or even negative correlations with stocks and bonds (Erb & Harvey, 2006) and may constitute a hedge against inflation (Bodie & Rosansky, 1980). However, these findings are not undisputed and some argue that the substantial growth leads to a "financialization" of commodity markets (Domanski & Heath, 2007). In this context, an increase in market participants, that simultaneously hold commodities and other financial assets, causes closer market integration and therefore undoes low asset return correlations and the resulting diversification potential (Tang & Xiong, 2012; Silvennoinen & Thorp, 2013; Lombardi & Ravazzolo, 2016; Ohashi & Okimoto, 2016).

Another related aspect of the controversial growth in commodity markets is the emergence of more convenient ways to trade them. In particular, futures, exchange-traded funds (ETFs) and exchange-traded notes (ETNs) are frequently used to track indices related to commodities. Index funds, in turn, invest into a range of commodity futures contracts and therefore represent an easy way to obtain broad exposure towards the whole commodity market or specific sectors. Aware of this fact and given the growing investor demand, the financial industry is constructing increasingly more sophisticated index products. Studying these innovations, Miffre (2012) labels and defines the *first*, *second* and *third generation* of commodity indices. While first generation indices seek to passively represent the whole commodity market, second generation indices rather mimic active investment strategies. The latest introduced, third generation indices, adopt these active approaches and additionally are short in commodity futures.

Although there is evidence on the profitability of tactical asset allocation in commodity markets (Miffre & Rallis, 2007; Rallis, Miffre, & Fuertes, 2013), the literature has so far paid surprisingly little attention towards enhanced commodity indices. In fact, only Daskalaki, Skiadopoulos, & Topaloglou (2017) and Kremer (2015) explicitly deal with third generation indices in a portfolio setting and confirm their superiority over the first and second generations. However, both studies consider only a small sample of commodity indices that may fail to adequately represent the variety of these investment vehicles. Motivated by this lack of empirical evidence and the simplicity to implement potentially profitable tactical asset allocation via enhanced indices, in this paper, we investigate whether investors can achieve improved portfolio performance by adding first, second or third generation commodity indices to their portfolios. We therefore contribute to the literature in several ways.

First, the potential diversification benefits of commodity indices are investigated by implementing a variety of mean–variance (mv) spanning tests. In this context, the whole sample and a specific sub-period of live-only index data are studied. Second, this evidence is extended by out-of-sample portfolio optimization that employs two different mv strategies, as well as $1/N$ with rebalancing ($1/N$), risk-parity (RP) and reward-to-risk timing (RRT). The resulting out-of-sample return time series with and without commodity investments are evaluated net of transaction costs and with multiple performance measures. Finally, the underlying sample consists of 21 commodity indices that cover first, second and third generation indices from different index providers and with distinct strategies. This dataset significantly expands existing research on enhanced commodity indices and provides more comprehensive and robust evidence.

The empirical findings for each of the three index generations can be summarized as follows. First generation commodity indices are not found to yield significant diversification benefits to equity–bond portfolios. Although they show low correlations with traditional assets, they have rather poor stand-alone investment properties, are spanned by equities and bonds when the full sample is considered and fail to improve out-of-sample portfolio Sharpe ratios (SRs). The performance of second generation commodity indices is superior to the first generation but the evidence is not clear-cut. On the one hand, they exhibit low return correlations, an equity-like stand-alone performance and consistent improvements of the out-of-sample performance. On the other hand, changes in portfolio SRs are not statistically significant and for most indices, the null of mv spanning cannot be rejected. Turning to the evidence of third generation commodity indices, their behavior is different from the first and second generation and also more heterogenous within the group. In fact, mv spanning is consistently rejected for these indices and therefore indicates a significant improvement of optimal portfolios on the augmented efficient frontier. However, these findings are only supported for half of the indices out-of-sample. The other half delivers worse out-of-sample performance than the benchmark. A potential explanation for these results stems from the short futures investments of the third generation commodity indices. They emphasize the individual active management return components of an index and simultaneously weaken the exposure towards changes in spot prices that are usually similar to all commodity indices.

Overall, while we can confirm the findings of Daskalaki et al. (2017) and Kremer (2015) that more recently developed commodity indices are, on average, superior to their predecessors and thus mimicking active investment strategies appears to be beneficial, we document that the evidence on out-of-sample portfolio performance for the third generation indices is less clear-cut than found in these studies.

Our work relates to previous studies on the performance of commodity investing. The early work of Bodie & Rosansky (1980) investigates an equally weighted fully-collateralized portfolio of up to 23 commodity futures over the 1950–1976 period. The authors show that commodities offered almost the same mean return and standard deviation as stocks over their sample period. Moreover, they find that commodity futures have higher positive skewness than stocks, offering the investor upside potential as opposed to downside risk. Similar results are presented by Gorton & Rouwenhorst (2006), who consider up to 36 commodity futures over an extended period from 1959 to 2004 and emphasize the equity-like investment properties of commodity futures. In another related study, Erb & Harvey (2006) first find equivalent risk–return properties of the S&P 500 Index and the S&P-GSCI in the 1969-2004 period. Subsequently, however, the authors point out that historical average geometric excess returns of commodity futures are close to zero. Erb & Harvey (2006) relate these seemingly inconsistent results to the "diversification return", which is "the difference between a portfolio's geometric return and the weighted-average geometric return of the portfolio's constituents." The authors state the magnitude of the diversification return as high as the risk premium found by Gorton & Rouwenhorst (2006), namely between 3% and 4.5%. To keep it short, most of the empirical literature reports inferior average stand-alone performance of individual commodity futures, portfolios thereof and commodity indices compared to equities and bonds for many time periods.[1] Nonetheless, attractive individual risk–return character-

[1] Among them are Anson (1999), Jensen, Johnson, & Mercer (2000), Jensen, Johnson, & Mercer (2002), Daskalaki & Skiadopoulos (2011), Belousova & Dorfleitner (2012) Büyükşahin & Robe (2014) and Daskalaki et al. (2017).

istics are not a necessary condition for assets to provide benefits in a portfolio setting.

The majority of the literature studying the profitability of commodities in a portfolio context performs different applications of spanning tests, first introduced by Huberman & Kandel (1987). Spanning tests base on the mv framework of Markowitz (1952), and investigate whether the introduction of an additional asset significantly enhances the efficient frontier. A rejection of the spanning hypothesis therefore indicates a statistically significant improvement of the investment opportunity set. Scherer & He (2008), Daskalaki & Skiadopoulos (2011), Belousova & Dorfleitner (2012) and Kremer (2015) study individual commodity futures, commodity indices, mv and non-mv consistent utility functions and different time periods. However, their results are mixed and the potential profitability of commodity investments remains unclear. The recent literature also extends the in-sample spanning tests with out-of-sample analyses. To begin with, Daskalaki & Skiadopoulos (2011) determine optimal portfolios by direct utility maximization and find no diversification benefits of commodities for different utility function. In contrast, Bessler & Wolff (2015) and Kremer (2015) present evidence in favor of commodity investments when implementing various trading strategies. Lastly, Daskalaki et al. (2017) apply the concept of stochastic dominance efficiency to test in-sample and out-of-sample portfolio gains of commodities and also find supportive results.

The remainder of this paper is structured as follows. Section 2 briefly reviews the market for commodity index investments. Section 3 presents the data and the empirical methodology. Section 4 presents and discusses the main results, while Section 5 provides several robustness checks. Section 6 concludes.

2 Commodity Index Investments

While there are several opportunities to invest into commodity markets, most of them come with drawbacks. First, holding physical commodities is undoubtedly impractical and expensive for an average investor. Second, stocks of commodity related firms are shown to be poor substitutes for commodity investments (Gorton & Rouwenhorst, 2006), as they contain a noisy component that stems from the equity nature of the investment (Geman, 2005). Third, managed futures programs come with high fees that usually amount to an annual 2% of the notional and an additional 20% of the profits (Stoll & Whaley, 2011). Fourth, trading commodity futures is the most popular approach but still requires frequent monitoring of margin accounts and rolling of contracts. Picking up the flexibility of futures and avoiding the steady need for investor activity, commodity index related investments likely represent the most convenient way to broadly invest into the commodity market (Geman, 2005). Besides trading commodity futures according to the index providers' handbook on one's own, several financial instruments allow simplified access to commodity indices (Tang & Xiong, 2012; Jensen & Mercer, 2011). In the case of a commodity index swap, the investor can obtain a variable leg that is tied to the price development of the index. Index swaps are primarily used by institutions and their OTC nature introduces additional counterparty risk. ETFs, on the other hand, are traded like stocks and usually replicate an index by trading the respective futures. Furthermore, ETNs can be used in a similar way to ETFs and achieve a closer tracking of the index due to contractual replication with third parties. However, this advantage is accompanied by additional counterparty risk.

Miffre (2012) divides today's spectrum of commodity indices into three index generations that have emerged over time. The *first generation* incorporates early introduced indices, whose common properties are long-only investments in liquid contracts at the front end of the term structure and rather infrequent rebalancing. Furthermore, they assume full-collateralization, i.e. that futures margins amount to 100% of the contracts' notional value (Miffre, 2012). Overall, first generation commodity indices try to passively represent the commodity market. For example, the S&P Goldman Sachs Commodity Index (S&P-GSCI) is constructed according to the amount of world production of each underlying commodity (S&P Dow Jones Indices, 2017). Erb & Harvey (2006) note that commodity indices primary represent commodity portfolio strategies, because "there is no agreed-upon way to define the composition of the aggregate commodity futures markets". In line with this interpretation, Miffre (2012) argues that first generation commodity indices implicitly assume a backwardated futures term structure as they are long-only invested. It follows that their performance, more precisely the roll return, tends to be good in backwardated but relatively poor in contangoed commodity markets.

This inherent tendency of the first generation indices to underperform in contangoed futures markets motivated the invention of the *second generation* commodity indices. While they are still long-only investments, these "enhanced" indices deviate from the goal of market representation. Instead, second generation indices apply different strategies and construction rules to earn higher risk-adjusted returns that are supposed to be independent of the shape of the futures term structure (Daskalaki & Skiadopoulos, 2011). Miffre (2012) differentiates between indices that alter futures rolling approaches in order to weaken the negative impact of rolling in contangoed markets and indices that apply more familiar investment strategies from the equity markets – for instance, momentum and mean reversion.

Finally, the *third generation* of commodity indices differentiate from the second generation by additionally allowing short positions. Miffre (2012) categorizes four different types of indices that are currently present on the market.[2] First, indices that follow a momentum strategy, whereby the historical price path is decisive for long or short investments. Second, indices that exploit the commodity futures term structure and aim at maximizing the implied roll return by going long in backwardated futures and short in contangoed commodity futures. Third, indices that enter into an equal amount of long and short positions at the same time in order to be "market neutral". The last category includes indices that derive trading signals from the combination of multiple factors, including technical and fundamental analysis. To sum up, there is a broad range of commodity indices that use several strategies with different implicit degrees of active management.

3 Data and Empirical Methodology

3.1 Commodity Indices

In order to examine whether commodity indices can enhance the performance of portfolios, a broad sample should be considered in order to ensure market representativeness. To this end, three first generation, nine second generation and nine third generation commodity indices are studied. In the following, we outline the reasons to include specific indices and provide short descriptions of each. More detailed information on methodology, constituents, base and launch date of each commodity index is provided in Table A.1 of the Appendix.

First of all, the well-known S&P Goldman Sachs Commodity Index (S&P-GSCI) and the Bloomberg Commodity Index (BCOM) are included for first generation commodity indices as they are most commonly used in empirical studies and are also of practical relevance.[3] Both, the S&P-GSCI and the BCOM aim at representing the broad commodity market. The weights of the former index are derived from the last five years' world production of the

[2] See also Miffre (2016) for a more recent review of long-short investments in commodity markets.
[3] According to Tang & Xiong (2012) and Stoll & Whaley (2011), both indices are the largest by market share.

underlying commodities (S&P Dow Jones Indices, 2017). Consequently, the S&P-GSCI has currently a high weight of 56.24% in energy commodities. The BCOM sets each commodity's weight based on measures of futures liquidity and commodity production. Additionally, it has implemented a diversification rule, defining lower and upper bounds on single commodity weights, within the index. Thus, the aggregated sector weights are more balanced, as indicated by a weight in energy commodities of only 30.57% (Bloomberg, 2017). The third first generation index we consider is the Deutsche Bank Liquid Commodity Index (DBLCI), which differs from the S&P-GSCI and BCOM in two aspects. On the one hand, the DBLCI has a more concentrated exposure to only six commodities with fixed and predefined weights. On the other hand, the DBLCI has a different roll schedule of commodity futures contracts with longer maturities (Deutsche Bank, 2008).

Our choice of second generation commodity indices takes several aspects into account. First, investigating different commodity indices from the same index provider may allow for interesting insights by direct comparison. We consider the Deutsche Bank Liquid Commodity Index Optimum Yield (DBLCI-OY) which is a second generation counterpart of the DBLCI. The DBLCI-OY includes the same six commodities as the DBLCI, but selects single commodity contracts from a greater range of maturities and based on their implied roll returns. Closely related to this index, the Deutsche Bank Liquid Commodity Index Optimum Yield Balanced (DBLCI-OYBA) and the Deutsche Bank Liquid Commodity Index Optimum Yield Broad (DBLCI-OYBR) follow the same selection scheme, although they differ with respect to their index constituents and weights. The last member of this index family, the Deutsche Bank Liquid Commodity Index Mean Reversion (DBLCI-MR), follows the same base weights as the DBLCI, but adjusts single commodity weights with the purpose of buying futures cheap and selling them high, thus assuming a long run mean reversion in commodity prices (Deutsche Bank, 2008). The second index family considered is the one from Morningstar, more precisely the Morningstar Long/Flat Commodity Index (MSLF) and Morningstar Long-Only Commodity Index (MSLO). The MSLF holds long positions in commodities with a positive price momentum signal and cash positions otherwise while the MSLO takes long positions in all eligible commodities of the MSLS (Morningstar, 2013).

In order to obtain meaningful insights into commodity indices as a group, a second aspect of index selection is to cover indices with a variety of strategies and constituents. For this reason, the SummerHaven Dynamic Commodity Index (SDCI) is considered. This equally weighted index combines signals of futures backwardation and price momentum, and invests in different contract maturities (SummerHaven Index Management, n.d.). Next, we consider the CYD Long Only Commodity Index (CYDLO) which invests in backwardated commodity futures and distinguishes itself by choosing from the whole futures term structure (CYD Research, 2013). Finally, the Merrill Lynch Commodity Index eXtra (MLCX) represents another second generation index that only slightly differs from the first generation by rolling second- into third-month contracts, instead of nearby contracts, over an extended 15-days roll-period (Merrill Lynch, 2006).

The selection of third generation indices is somewhat limited by data availability as they are new developments that have emerged since the late 2000s. Nonetheless, nine commodity indices of the third generation are subject of this study. The Morningstar Long/Short Commodity Index (MSLS) employs a momentum rule that relates a linked price series of each commodity to its 12-month moving average price. The resulting signal determines each commodity's weight in the index and whether it enters with a long or short position. The Morningstar Short-Only Commodity Index (MSSO) is short in all constituents of the MSLS whereas the Morningstar Short/Flat Commodity Index (MSSF) holds the same short positions as the MSLS but replaces long futures by cash investments (Morningstar, 2013). The Morningstar index family shares the same commodity universe and underlying methodology as well as index base and launch dates. These properties enhance conclusive inter-generation comparisons of commodity indices.

We also consider the Credit Suisse Momentum and Volatility Enhanced Return Strategy (CSMOVERS) and the Credit Suisse Momentum and Volatility Enhanced Return Strategy Market Neutral (CSMOVERSMN). The CSMOVERS invests equally weighted in ten out of 24 eligible S&P-GSCI sub-indices and exploits the volatility signals of the underlyings to choose contract maturities of either one, three or six months. Subsequently, the price momentum of each commodity is decisive for either long or short positions. The market neutral version of the index, CSMOVERSMN, applies the same methodology, but is required to simultaneously hold six long and six short commodity futures (Credit Suisse, 2010). Three third generation commodity indices from CYD make up another related index family. First, the CYD Long–Short Commodity Index (CYDLS) and the CYD Diversified Long/Short Commodity Index (CYDDLS) both follow term structure strategies. While the implementation of the CYDLS is largely unconstrained, the CYDDLS requires a more balanced ratio of long and short positions as well as minimum commodity sector weights (CYD Research, 2013; Vescore Indices, 2016). Second, the CYD Market Neutral Plus Commodity Index (CYDMNP) is included, and follows an interesting approach that invests in multiple contracts of the same commodity at the same time. In fact, each constituent is represented by a short position in the nearby contract and, based on liquidity, by up to two long positions in the 2nd- and 3rd-nearby contracts (CYD Research, 2013).[4] The CYDMNP is the only leveraged investment considered in this study. With a leverage factor of two, the index enters futures positions with the notional being twice as big as the invested money. Leveraged investments in commodity markets are usually considered to be easily achievable. However, the leveraged nature of the index affects its returns and volatility and has to be taken into account when comparing it to non-leveraged investments. Nonetheless, the index is included due to its particular strategy and the practical relevance of leveraged commodity exposure. Finally, the Barclays Backwardation Long/Short Index (BBLS) represents the last third generation commodity index. It is long in the three-month maturity contracts of the six commodities with the greatest backwardation and short in the nearby contracts of the six commodities with the least backwardation (Barclays, 2014).

3.2 Data

The commodity index time series considered are based on total return calculations. Moreover, data on assets that represent the traditional investment universe are required. To this end, the S&P 500 Total Return Index (S&P 500) and the Bloomberg Barclays US Aggregate Bond Index[5] (Barclays Bonds) are used to proxy diversified passive stock and bond investments, respectively. To account for dynamic equity market portfolios, returns on the Fama & French (1993) size (Small minus Big, SMB) and value (High minus Low, HML) portfolios are considered. As risk-free rate, returns of the 1-month U.S. Treasury Bill are used. The dataset is denominated in U.S. dollars, has monthly frequency and covers the period from February 2000 to April 2017, i.e. 207 simple return observations are computed for each time

[4]Simultaneously holding long and short positions with different maturities in the same commodity is also referred to as holding "spreading" positions (Szymanowska, DeRoon, Nijman, & Goorbergh, 2014).

[5]This index covers Treasuries, government related and corporate securities, mortgage-backed securities, asset-backed securities and commercial mortgage-backed securities denominated in U.S. dollars (Bloomberg Barclays Index, 2017).

series.[6] Commodity, stock and bond data primarily come from Thomson Reuters Datastream and are complemented by data from Bloomberg and index providers' websites. The SMB and HML portfolio returns, as well as the risk-free rate, are obtained from Kenneth French's data library.[7]

3.3 Mean–Variance Spanning Tests

Spanning tests are frequently used in the literature to examine the diversification benefits of adding an asset to an existing portfolio. The underlying methodology was first introduced by Huberman & Kandel (1987) and compares an initial asset universe, consisting of some benchmark assets, with an augmented asset universe, containing the same benchmark assets plus some test assets. More precisely, the mv frontiers of both asset universes are studied under the assumption that investors derive utility only from the mean and variance of returns. As proposed by Huberman & Kandel (1987), *spanning* is given if both frontiers coincide and thus there is no enhancement of the investment opportunity set, i.e. investors do not benefit from adding the test assets to the benchmark portfolio. We first consider three variants of spanning tests, namely the Wald, the Lagrange Multiplier (LM) and the Likelihood Ratio (LR) test. For each test, the null hypothesis states that the test assets are spanned by the benchmark assets. The Technical Appendix provides more details and formulas for implementing these tests.

Kan & Zhou (2012) propose several extensions to the spanning literature to address some of their shortcomings. First they suggest a step-down test that allows to disentangle differences due to return enhancement on the one side and risk reduction on the other. As outlined above, mv spanning tests examine the statistical significance of changes in the efficient frontiers of the initial and augmented asset universes after test assets are included. Merton (1972) shows that an efficient frontier is determined by the location of the Global Minimum Variance Portfolio (GMVP) and the Tangency Portfolio (TP). Based on this fact, Kan & Zhou (2012) point out that the spanning hypothesis can be understood as a joint test of changes in the GMVP and the TP. In this context, the authors show that spanning tests put relatively more weight on potential changes of the GMVP than they do for changes in the TP. Consequently, span-

[6] Belousova & Dorfleitner (2012) note that the choice of simple or logarithmic returns is not trivial in the context of spanning tests. However, simple returns are preferred in this case, because portfolio aspects are of primary interest. Simple returns are also used by Daskalaki & Skiadopoulos (2011).

[7] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html.

ning tests may fail to reveal slight changes in the TP (return enhancements) that, however, are economically significant. At the same time, they may overvalue small changes in the GMVP (risk reduction). The step-down test of Kan & Zhou (2012) deals with this limitation, as it disentangles the two effects. The so-called F_1 test considers the null of no return enhancement while the F_2 test considers the null of no risk reduction. We provide details on the step-down procedure in the Technical Appendix.

Another issue of standard spanning tests implemented by maximum likelihood estimation is the assumption of normally distributed asset returns. Kan & Zhou (2012) show that non-normality of returns does not affect the asymptotic distributions and only marginally alters the finite sample distributions of the spanning tests, given that the error term is conditionally homoskedastic. However, in case it is conditionally heteroskedastic and returns are non-normal, the authors show that the usual test statistics are no longer chi-squared distributed under the null hypothesis and thus over- or underrejection of spanning can result. For this purpose, Kan & Zhou (2012) present a spanning test based on the generalized method of moments (GMM) methodology of Hansen (1982) that is not dependent on the normality assumption for asset returns. Again, details are provided in the Technical Appendix.

3.4 Portfolio Optimization

This section describes the methodology used to investigate the potential out-of-sample benefits of commodity indices within diversified investment portfolios. To this end, we adopt the rolling sample approach of DeMiguel, Garlappi, & Uppal (2009) for a range of asset allocation strategies. An estimation window of M months past return observations is used to estimate the parameters, such as expected returns and the covariance matrix, for month t, starting with $t = M + 1$. These parameters first allow us to determine optimal portfolio weights for each asset allocation strategy and then calculate the respective out-of-sample returns for month t. In $t + 1$, the estimation window is rolled forward by one month by dropping the oldest and adding the most recent monthly return observation. This procedure is then repeated for the whole sample and results in $T - M$ return observations for each strategy. To ensure robust results, the rolling sample is implemented for different estimation windows of M = 36, 48 and 60 months for the period from February 2000 to April 2017.

In this process, multiple asset allocation strategies are applied: mean–variance (mv), naive diversification with rebalancing ($1/N$), risk-parity (RP) and reward-to-risk timing (RRT). Details on the respective strategy and their implementation are provided in the Technical Appendix.

Initially, each of the asset allocation strategies is implemented for a benchmark portfolio consisting of Equities (S&P 500), Bonds (Barclays Bond Index), and the factor portfolios SMB and HML.[8] Afterwards, the available benchmark assets are augmented by one of the 21 commodity indices at a time and, again, optimal portfolios and their corresponding out-of-sample returns are calculated. At each of the $T-M$ rebalancing points, monthly returns are adjusted by proportional transaction costs that are assumed to be 30 basis points of the total transaction volume.[9]

To evaluate the impact commodities have on the portfolio performance, several performance measures are used. First, the annualized time series mean and volatility of net returns are reported. Second, the SR is computed as the mean of net excess returns divided by the volatility of net excess returns, both after transaction costs and over the risk-free rate. In order to test whether differences in the SRs of the benchmark and augmented portfolios are significant, the approach of Jobson & Korkie (1981), corrected by Memmel (2003), is used. Third, the portfolio turnover is calculated for each strategy i as the average sum of absolute changes in portfolio weights ω over each of the $T-M$ rebalancing points and across all N available assets (DeMiguel et al., 2009). This is formalized as follows:

$$\text{Portfolio Turnover}_i = \frac{1}{T-M}\sum_{t=1}^{T-M}\sum_{j=1}^{N}\left(|\omega_{i,j,t+1}-\omega_{i,k,t^+}|\right), \tag{1}$$

where $\omega_{i,j,t}$ is the portfolio weight of asset j at time t for strategy i, ω_{i,j,t^+} is the portfolio weight prior to rebalancing in $t+1$, i.e. the weight of time t altered by the asset returns from t to $t+1$, and $\omega_{i,j,t+1}$ is the portfolio weight after rebalancing at $t+1$. Note that $\omega_{i,j,t}$ generally deviates from ω_{i,j,t^+} due to realized asset returns.[10] Lastly, the average portfolio

[8] Note that each asset allocation strategy considered does not allow short selling, yet the SMB and HML portfolios implicitly do. Nonetheless, they are used to proxy an actively managed equity benchmark for third generation commodity indices that similarly invest in short commodity futures contracts.

[9] Alternatively, transaction costs of 50 basis points are considered in order to test the robustness of the results regarding this assumption. See DeMiguel et al. (2009) and Bessler & Wolff (2015) for similar choices of transaction costs.

[10] Similarly, $\sum_{j=1}^{N}\left(|\omega_{i,j,t+1}-\omega_{i,k,t^+}|\right)$ is used to determine the transaction costs in each month t of the rolling sample. Furthermore, note that the common difference between $\omega_{i,j,t}$ and ω_{i,j,t^+} causes the $1/N$ strategy to have a turnover different from zero.

weight of the commodity indices over the $T - M$ rebalancing points is reported for each strategy i, which is given by:

$$\text{Average commodity weight}_i = \frac{1}{T-M} \sum_{t=1}^{T-M} \omega_{i,t}^{com}, \tag{2}$$

where $\omega_{i,t}^{com}$ is the portfolio weight of commodities at time t for strategy i. This measure helps to evaluate the extent to which commodities impact the results of each asset allocation strategy. Furthermore, it gives an idea of the optimal relative allocation that should be made towards commodities.

4 Results and Discussion

4.1 Descriptive Statistics and Correlation Analysis

Table 1 presents the descriptive statistics of all the assets that are used throughout the empirical investigation. To begin with, first generation commodity indices show, on average, low annualized returns of 2.41%, the highest annualized volatilities of 20.11% and consequently the smallest SRs of 0.04. Therefore, first generation commodity indices do not seem to represent good stand-alone investments. Next, second generation commodity indices have, on average, substantially higher returns (7.07%) than first generation indices and also exceed those of the S&P 500 index (6.15%). Their average annualized volatility of 17.21% is below first generation commodity indices and above that of equities. Overall, descriptive statistics point out similar stand-alone qualities of second generation commodity indices and equities, while even having a slightly higher SR of 0.33 (S&P 500: 0.31). Both first and second generation commodity indices are negatively skewed and exhibit excess kurtosis, which suggests more downside risk and fatter tails compared to normally distributed asset returns. This non-normality is confirmed by the Jarque–Bera test. Likewise, the last column of Table 1 reports the p-values of the ARCH test of Engle (1982). Finally, third generation commodity indices have, on average, similar returns (6.55%) as the S&P 500 and second generation indices, but are considerably less volatile (10.38%). Thus, their average SR of 0.56 is the second highest of the sample, below the bond index, which is the best performing asset over the period considered with a SR of 1.05. Contrary to first and second generation indices, the third generation shows positively skewed asset returns and therefore rather upside than downside risk.

Although descriptive statistics of the first and second generation indices are rather similar, there is more variety within the group of third generation indices. First, BBLS, CYDDLS and CYDMNP are outperforming the average commodity index by far in terms of the SRs of $0.97, 1.25$ and 1.29, respectively. While the former two do not show large methodological differences to the remaining indices, CYDMNP stands out because of its leverage factor and unique investment strategy. Nonetheless, a leverage should, in theory, not substantially affect the SR and therefore its strategy may be the source of outperformance.[11] Second, MSSF and MSSO both clearly underperform the other assets with annualized returns of 1.12% and -1.53% that are below the risk-free rate of 1.59%. Both indices fully exclude long commodity investments, which differentiates them from other indices. Overall, the average second and third generation commodity indices represent an attractive stand-alone investment with arguably superior properties than equities over the time period studied. However, significant differences exist within the group of third generation commodity indices.

The most relevant aspect for potential diversification gains of commodities is their correlation with traditional asset classes. To this end, Table 2 reports the pairwise correlations of all commodity indices and benchmark assets considered. To begin with, first and second generation commodity indices are typically highly correlated with other indices in their respective group and across groups, with correlations ranging from 0.74 to 0.98. Likewise, the two index generations are similarly correlated with the benchmark assets. Both are largely uncorrelated with bonds, style portfolios and the risk-free rate and, except for the MSLF, they are significantly positive correlated with the S&P 500 in a range of 0.28 to 0.37. These patterns suggest that both index types have the opportunity to diversify traditional portfolios but also that they are rather similar investments. This is, after all, as expected, because both index generations are long-only invested in a relatively small commodity investment universe. Next, third generation commodity indices behave differently in several aspects. First, their correlations with first and second generation indices is non-uniform, ranging from -0.94 to 0.53 and -0.96 to 0.88, respectively. Second, the correlations with the S&P 500 are either insignificant or negative, ranging from -0.41 to 0.01, and thus the reverse of first and second generation indices. A plausible explanation for both findings is the augmented commodity

[11] Paschke, Prokopczuk, & Wese Simen (2017) study a so-called “Curve Momentum” strategy that is closely related to that of the CYDMNP. The authors find similar risk–return properties for a diversified portfolio.

investment universe of third generation indices, i.e. the inclusion of short futures positions. Given positive correlations within different long commodity futures returns and between long futures and the S&P 500, it is reasonable to find opposite results when short futures are considered. The high negative correlation of the MSSF (-0.41) and MSSO (-0.32), which both fully exclude long futures positions, are further in line with this argument. Lastly, third generation commodity indices are also more positively correlated with the risk-free rate. Surprisingly, they do not show higher correlation with the style portfolios than first and second generation indices, although both represent an active investment approach. Overall, return correlations reveal that the third generation of commodity indices clearly differs from previous generations and may even better suit the diversification task.

To summarize, first and second generation commodity indices have highly correlated returns but differ with respect to their realized performance. While second generation indices are already an attractive stand-alone investment, first generation indices fall behind. However, both groups show low or zero correlations with traditional assets. Consequently, first generation commodity indices may still be valuable additions in a diversified stock–bond portfolio. Last but not least, third generation indices have, on average, superior SRs and even more promising correlations with traditional assets. Therefore, the analysis of descriptive statistics and correlations suggest the diversification potential of second and third generation indices and is inconclusive towards the portfolio gains of first generation indices. This evidence is consistent with previous research but is at the same time insufficient to answer the question whether investing in commodity indices is beneficial for an investor.

4.2 Mean–Variance Spanning Tests

The null hypothesis of spanning states that there are no statistically significant differences between the efficient frontiers of an initial and an augmented asset universe. This section discusses the results of several approaches that test this hypothesis which we have introduced before (see the Techincal Appendix for details). Table 3 reports the results of this analysis, where each of the tests – Wald, LR, LM, GMM-Wald (heteroscedasticity-robust), F_1 and F_2 (step-down approach) – is applied for the case of two benchmark assets – Barclays Bonds and the S&P 500 – augmented by one out of 21 commodity indices at a time.

Evidence for the first generation commodity indices shows that all of them fail to reject the joint null hypothesis of spanning for each test implemented, except for a weakly significant risk reduction of the S&P-GSCI, as indicated by the F_2 test. These findings are largely in line with the expectations derived from the analysis of the descriptive statistics: the individual risk–return profile of first generation indices is insufficient to enhance portfolio returns but their low correlations can sometimes diversify the portfolio risks. Similar results for first generation commodity indices and mv investors have been reported, for example, by Daskalaki & Skiadopoulos (2011) or Kremer (2015), and rather mixed evidence is presented by Scherer & He (2008) or Huang & Zhong (2013).

The results for the second generation commodity indices yield similar conclusions. Except for the MSLF, the spanning hypothesis of the Wald, LR, LM and GMM-Wald tests cannot be rejected for any of the indices. Looking at the F_1 and F_2 tests, weak evidence exists for a return enhancement of the SDCI and a risk reduction of the DBLCI-OY, both at the 10% significance level. Nonetheless, the average second generation commodity index fails to significantly improve the augmented efficient frontier. These results are unexpected, given the indices' descriptive statistics and the results of Kremer (2015) and Daskalaki et al. (2017). Note, however, that both studies use a different time period and fewer indices. Moreover, results for one of their indices, the MSLF, are similar to the findings outlined in Table 3.

The results change for the third generation commodity indices, as the evidence indicates their diversification potential. Including each of the nine indices leads to a highly significant rejection of the mv spanning hypothesis, even when non-normality and conditional heteroscedasticity of residuals is accounted for by the GMM-Wald test. The step-down procedure further reveals the sources of diversification gains. Accordingly, the F_2 test shows that each index is able to reduce the overall portfolio risk, and at the same time, the F_1 test suggests that five out of nine indices can enhance the portfolio return at the 1% significance level. As a whole, the spanning test results indicate an improvement of portfolios by the inclusion of third generation commodity indices that is primary driven by a risk reduction.

4.3 Portfolio Optimization

As a first step, we implement an in-sample portfolio optimization to obtain numerical results of the maximum achievable diversification benefits from commodity indices. At this stage, it is sufficient to consider the two mv strategies that, naturally, dominate other asset allocation approaches in-sample. Panel A and B of Table 4 report these results for the aggressive (low risk aversion) and conservative (high risk aversion) investor types, respectively. First, in line with the respective setups, the aggressive mv strategy shows overall higher returns, volatilities, SRs and portfolio turnovers than the conservative approach.[12] These different findings trace back to the more strict portfolio constraints imposed on the conservative mv strategy that consequently limits the exposure to risky assets. Second, adding commodities within each of the in-sample mv strategies substantially enhances the portfolio performance. Independent of the commodity index included, SRs consistently improve and the majority of these changes is highly statistically significant. More precisely, the initial aggressive and conservative benchmark SRs amount to 4.32 and 4.09 and improve towards values within the range of 4.84 to 5.75 and 4.67 to 5.42, respectively. Qualitatively similar in-sample results of adding commodity indices are reported by Bessler & Wolff (2015), yet their returns and SRs are smaller compared to the results given in Table 4. One possible reason for this performance gap is the different time period studied. More likely, however, the primary reason is the greater benchmark asset universe we consider compared to Bessler & Wolff (2015), who do not include dynamic value and size equity portfolios. A third interesting aspect are the considerable differences in the commodity weights of the aggressive and conservative strategies. Averaging over all commodity indices, the aggressive approach allocates 24.55% and the conservative one 11.78% towards commodities. This observation suggests that commodities rather serve as risky and return enhancing assets than low-risk portfolio diversifiers.

Overall, and not surprisingly, each generation of commodity indices can enhance the portfolio performance in-sample. However, in-sample results are not applicable to real world asset management. Consequently, we now focus on the results of the out-of-sample portfolio optimization. This approach looks at the practically achievable impact commodities could have had on an equity–bond portfolio over the time period studied. Table 5 reports the out-of-sample results derived for an estimation window of 36 months and transaction costs of 30 basis points.

To begin with, the addition of first generation commodity indices to the benchmark assets only slightly impacts the out-of-sample performances of each strategy. Most of the time,

[12] Note that the turnovers of both strategies are substantial. For instance, a turnover of 100%, as defined earlier, means that an investor reallocates half of his portfolio in each month. This is, however, expected in the in-sample mv case. At each rebalancing point, the wealth is allocated towards the asset with the highest return in the next period, while still complying with the portfolio constraints.

returns and volatilities increase simultaneously and SRs maintain a similar level. For instance, the mv (aggressive) strategy has the highest relative allocation towards first generation commodity indices, which ranges from 14.38% to 26.53%, and still almost identical out-of-sample SRs for the benchmark and augmented case. In contrast, the mv (conservative) approach allocates merely between 4.04% and 7.03% in commodities and realizes an increase in the SRs of approximately 0.1. Moreover, the SRs of RP and RRT strategies are largely unchanged and those of the $1/N$ are noticeably lower. Similarly, the portfolio turnover does not materially change for each of the strategies considered. In short, first generation commodity indices fail to consistently improve out-of-sample portfolio performance in the period studied. Therefore, the previously reported in-sample portfolio benefits of these indices do not remain out-of-sample. Beyond that, there are only marginal differences between the inclusion of the BCOM, DBLCI or S&P-GSCI, i.e. they serve as homogenous investment tools, in line with their high pairwise correlations. Qualitatively similar results are reported by Daskalaki & Skiadopoulos (2011) and Kremer (2015). Bessler & Wolff (2015) report mixed results for changes in SRs when investing in the S&P-GSCI.

Turning to the inclusion of second generation commodity indices, out-of-sample returns are usually above the benchmark and first generation index cases. At the same time, return volatilities are higher relative to the benchmark but comparable to the first generation. In total, these changes increase SRs notably, except for the $1/N$ strategy. For instance, the benchmark SR of the mv (conservative) strategy increases from 0.79 to values ranging from 0.88 to 1.17. Nonetheless, only some changes in SRs are found to be statistically significant, namely in case of the DBLCI-OYBA and SDCI. Note that, consistently for every strategy, the SDCI exceeds the portfolio gains of every other second generation commodity index. On the one hand, this may be caused by its particular methodology that combines signals of futures term structure and price momentum. On the other hand, the SDCI makes more use of backfilled data than other second generation commodity indices.[13] Concerning the portfolio turnover, including second generation commodity indices does not lead to noteworthy changes. In contrast, the average commodity index weight is considerably higher compared to the setting with first generation indices. For example, the average commodity allocation of the RRT strategy almost doubles from 3.13% to 5.98% when the second generation is available. In total, the portfolio impact of different second generation indices is

[13]We discuss the issue of backfilled data in the next section.

quite homogenous, although not to the same extent as the first generation. Finally, the previously reported in-sample diversification benefits from second generation indices are partially preserved and an investor is usually better off including one of them.

Concerning third generation commodity indices, the results are more heterogenous. In fact, two sub-groups of indices can be formed based on their portfolio impact. First, BBLS, CSMOVERS, CSMOVERSMN, CYDDLS and CYDMNP consistently enhance the out-of-sample performance relative to the benchmark case. More specifically, they improve portfolio returns for each strategy and sometimes additionally reduce return volatilities. Therefore, SRs are above the benchmark case for every strategy as soon as one of these indices widens the investment opportunities. Additionally, the changes in SRs are statistically significant for BBLS, CYDDLS and CYDMNP in each strategy and in one out of five strategies for the CSMOVERS and CSMOVERSMN. Reported SR improvements further exceed those of first and second generation commodity indices. For instance, the benchmark SR of the mv (aggressive) strategy rises from 0.68 to values within the range of 0.90 to 1.54, whereas the best second generation index achieves an SR of 1.03. Another finding is that average commodity weights are higher relative to first and second generation indices. Within the mv (conservative) strategy, the five outperforming third generation indices represent, on average, 43.48% of the portfolio allocation. It is also worth mentioning that, within the RP and RRT strategies, the portfolio weights of the CYDMNP amount to 49.84% and 43.26% and thus by far exceed those of other commodity indices. These allocations trace back to the low return volatility of the index, which is similar to that of the Barclays Bond Index. Lastly, the inclusion of third generation commodity indices does not notably change the portfolio turnover relative to the other cases. Overall, BBLS, CSMOVERS, CSMOVERSMN, CYDDLS and CYDMNP realize significant out-of-sample diversification benefits, when added to a stock–bond portfolio, that are beyond those of second generation commodity indices.

In contrast to these findings, a second group of third generation commodity indices, namely the CYDLS, MSLS, MSSF and MSSO, clearly underperforms both the benchmark case and the other commodity indices. Returns of portfolios including these indices are below the benchmark case, and they also fail to reduce risks. As a consequence, out-of-sample SRs are usually lower than their benchmark portfolios. At the same time, the portfolio turnover increases for both mv strategies relative to other indices. This fact is one possible reason for the underperformance within these strategies, although it does not apply to the $1/N$, RP and

RRT strategies. Turning to their average portfolio allocation, no special pattern is apparent besides relatively low weights in the mv (aggressive) strategy. In total, these four commodity indices fail to maintain their in-sample diversification benefits and partly worsen the out-of-sample performance of a stock–bond portfolio. Contrary to these findings, Kremer (2015) reports improved portfolio performance from adding the CYDLS and MSLS. However, the author does not apply multiple asset allocation strategies, uses fewer benchmark assets and considers the 1991 to 2013 period that covers more backfilled data.

Although each third generation commodity index is consistently beneficial in-sample, the opposing out-of-sample performances are actually in line with the respective index SRs as outlined in Section 4.1. A first plausible reason for the more heterogenous and extreme performances within the group of third generation indices and compared to other commodity indices lies in their primary difference: the inclusion of short futures positions. On the one hand, third generation indices can establish similar long investments as second generation indices. On the other hand, they intensify their exposure to the pursued strategy by entering short positions in futures with reversed signals.[14] In other words, the return component, that is attributed to implicit active management, is more pronounced for third than for second generation indices. At the same time, simultaneously established long and short futures positions may reduce the overall exposure towards spot price movements on the commodity market, i.e. it weakens the influence of spot returns. Assuming an unexpected rise in the overall commodity market, long futures will generally increase in value while short commodity futures rather decrease in value. In case an index holds both long and short futures, it is reasonable that parts of the price movements cancel out. Taken together, third generation commodity indices may perform differently because they put more emphasis on the active return component that is unique for each index and less focus on the passive spot price movements that all commodity indices have largely in common. This argument may also explain the low volatilities of third generation indices reported earlier.[15]

Another explanation for the heterogenous performance within third generation indices stems from the explicit way they implement their strategies. Although over- and underperforming

[14] For example, second and third generation term structure indices can both invest long in backwardated commodity futures. However, the third generation index also obtains short futures contracts of commodities that are in contango and thus the index is relatively more committed to the term structure strategy.

[15] For instance, the CSMOVERSMN and CSMOVERS follow the same strategy but the former is more balanced with respect to long and short futures positions. In line with the argument, the CSMOVERSMN is considerably less volatile than the CSMOVERS (10.30% vs. 14.77%).

indices do not follow essentially different strategies, some detailed conclusions can be put forward. First, the underperforming CYDLS lacks a diversification rule and thus is prone to exclusively invest long or short in commodity futures. In contrast, the later introduced CYDDLS follows a similar term structure strategy, yet substantially outperforms the CYDLS. Comparing their methodologies, it becomes clear that the CYDDLS is designed to aim for a more balanced long–short exposure. Second, it is striking that each Morningstar third generation commodity index is underperforming. However, the three indices are based on the same price momentum rule, and if this approach fails in practice each of the Morningstar indices is likely to perform poorly. In line with this argument, the MSLF and MSLO also slightly underperform the average second generation commodity index. Nonetheless, a momentum rule itself is not necessarily the reason for bad performance. Both Credit Suisse indices, the CSMOVERS and CSMOVERSMN, also exploit momentum signals but successfully improve the out-of-sample portfolio performance. In turn, these two indices additionally take into account the commodity futures volatility to determine contract maturities, and this deviation may be the reason for significant differences in their performances. A third aspect that may explain the performance gap within third generation indices is the difference in index launch dates, i.e. a potential bias in backfilled data that is discussed next. In fact, the average overperforming third generation index has been launched several years after the average underperforming index, and thus they apply more optimized data.

5 Further Analyses

5.1 Backfilled Data

An important aspect we need to consider in order to check the robustness of our results is the issue of backfilled commodity index data, i.e. hypothetical index levels for a time period prior to the actual market launch of an index, which are commonly published by index providers.[16] While these data simulate the actual index performance in historical periods and thus are useful to get an idea of its performance, one has to be cautious when employing them. As Scherer & He (2008) note, index providers may optimize their products' methodologies based on historical market data, which, at least to some extent, support marketing. However, once an index is launched and "goes out-of-sample", its performance may be dif-

[16]Columns four and five of Table A.1 list the base and launch dates of each commodity index considered.

ferent, potentially worse than better, compared to the historical optimization. Some visual indication for this kind of bias in commodity index data are given by Figure 1, which illustrates the historical index levels of the BBLS and CYDMNP and highlights the dates of their market launches. Both panels of the figure suggest that in the years following the index launch, its performance seems to be worse than previously reported for the backfilled data. Note, however, that this pattern is not apparent for each commodity index, is not tested statistically and can ultimately have different reasons. Nonetheless, it is plausible to examine whether results are robust to this potential bias. With regard to the index sample studied, the time series of third generation indices consist of relatively more backfilled data than those of first and second generation indices. Therefore, in case there is a bias in the backfilled index data, it is more pronounced for third generation indices and may overestimate their performance. To account for this issue, we conduct a sub-sample analysis of live-only index data. In this analysis, the CYDDLS will be excluded for two reasons. First, the index implementation date is vaguely published as “Spring 2013”, thus no exact date is available. Second, apart from the CYDDLS, the most recently introduced index, the BBLS, was launched in November 2010. Therefore, excluding the CYDDLS adds more than two years of observations to the sub-sample analysis.

Table A.2 in the Appendix presents the sub-sample (2010-2017) results of the spanning tests.[17] We can observe that the spanning hypothesis is now rejected for each first and second generation index, regardless of the tests applied and at least at the 5% significance level. The majority of these indices simultaneously increase returns and decrease risks, relative to the initial situation, as indicated by the F_1 and F_2 tests. Spanning is still rejected for each third generation commodity index but the F_1 test reveals that portfolio returns are no longer enhanced. Instead, rejection of the null is caused by a significant risk reduction in the GMVP.

5.2 Time-Variation in Spanning

The sub-sample findings for first and second generation indices have some important implications, namely that results of spanning tests appear to be unstable over time. In order to further investigate this issue, the sample is arbitrarily split into four equally sized sub-

[17] Please note that we only conduct the spanning tests here, as the time series of available data is too short for the portfolio optimization analysis. If anything, this biases our study towards concluding that commodity index investing is beneficial for investors. However, our overall conclusion is that this is not necessarily the case.

samples and the tests are repeated.[18] Table A.3 in the Appendix reports the results for the Wald test and also lists the total number of rejections for each commodity index. The outcomes suggest that rejections of the spanning hypotheses are indeed dependent on the time period studied. In particular, first and second generation indices are shown to significantly improve the efficient frontier in at least one of the four periods. Nonetheless, third generation indices still reject spanning most of the time, although not consistently in every period and the results are also less significant. The fact that the results indicate some time-variation reveals a drawback of the spanning methodology: it implicitly assumes that the efficient frontier is constant over time, and hence also the optimal portfolios. If, however, expected asset returns and (co)variances are time-varying, parameter estimates and the efficient frontiers are not the same over all points in time as well. Consequently, spanning test results may not necessarily be informative for each sub-sample. In other words, an investor may realize diversification benefits from first or second generation commodity indices in certain sub-periods, although this is not revealed by spanning tests. This issue is better taken into account within the portfolio optimization approaches that estimate parameters on a monthly basis, and rebalance accordingly.

5.3 Spanning with More Benchmark Assets

The main results of mv spanning suggest that investors can realize diversification benefits from the inclusion of a third generation commodity index but not via first or second generation indices. Nonetheless, an important aspect of third generation commodity indices needs to be considered to obtain more robust results. This type of index serves as an active investment that incorporates long and short positions based on signals observed in the markets. Contrary, the two benchmark assets, the S&P 500 and Barclays Bonds, are both long-only investments that aim at passively representing the equity and bond market, respectively. Consequently, they may be inappropriate benchmarks for third generation commodity indices and more dynamic alternatives should be used. Therefore, following Daskalaki et al. (2017), returns of the Fama & French (1993) SMB and HML factor portfolios are added to the existing benchmark assets.

To begin with, Table A.4 in the Appendix shows that, naturally, test statistics decrease due

[18]The choice of four equally sized sub-samples is supposed to ensure a sufficiently large number of monthly observations. More precisely, the first sub-sample is based on 51 monthly returns and the other three on 52 observations, respectively.

to the inclusion of value and size equity portfolios in the benchmark assets. Nonetheless, the results are qualitatively unaffected, as only the significance of a few F_1 and F_2 test results of second and third generation indices have altered slightly. Therefore, the diversification benefits of third generation commodity indices do not vanish if the investor has access to dynamic equity investments.

Table A.5 in the Appendix reports the results of the 2010–2017 sub-sample analysis with an extended benchmark asset universe. Again, the test statistics are universally lowered compared to the previous case but the joint spanning hypotheses are still rejected for each commodity index at least at the 10% significance level. Noteworthy changes only appear for second generation indices that no longer diversify portfolio risk, as suggested by the F_2 test. In total, the sub-sample analyses reveal that first and second generation commodity indices offer diversification benefits in the more recent period. At the same time, evidence for third generation indices is slightly weaker than previously reported, but still the use of live-only data does not qualitatively change the results.

5.4 Different Transaction Costs and Estimation Windows

In order to test the robustness of the out-of-sample portfolio optimization results with respect to transaction costs, we repeat the analysis for transaction costs of 50 basis points instead of 30 basis points assumed before. The results are presented in Table A.6 in the Appendix. Naturally, this change affects asset allocation strategies with a high turnover – i.e. both mv strategies – more than those with low turnover – for instance, the $1/N$ strategy. Accordingly, out-of-sample returns decrease proportionally to the portfolio turnover for the benchmark and augmented scenarios. Due to similar turnover levels within each strategy and across different commodity indices, the relative performances are largely unchanged. As a consequence, the previously outlined conclusions remain and the overall results are robust to the choice of transaction costs.

As additional robustness tests, Tables A.7 and A.8 in the Appendix present the results when longer estimation windows of 48 and 60 months, respectively, are used to estimate model parameters. A first striking result is that the returns of the benchmark portfolios decrease consistently for each strategy once longer estimation windows are used. At the same time, the return volatility remains on a similar level and thus out-of-sample SRs are considerably lower. The same happens for the inclusion of first and second generation commodity indices. Moreover, these changes show the same magnitude and consequently the previously reported results are qualitatively unchanged. This overall tendency of worse performances for longer estimation windows signals that new information should be weighted relatively more strongly than earlier information. In other words, the applied asset allocation models incorporate these news too slowly. Another simple explanation for the decline in performance could be above average asset performance in those years that are now part of the estimation window and were previously used to determine the out-of-sample returns.[19] Looking at third generation commodity indices, mixed results are revealed. The earlier defined group of outperforming indices is still above average and in some cases their SRs are even enhanced for the extended estimation windows. Combined with the reduced SRs of the benchmark portfolios, the relative SR improvements of these commodity indices are larger and the changes are also more statistically significant. As a consequence, out-of-sample portfolio benefits of BBLS, CSMOVERS, CSMOVERSMN, CYDDLS and CYDMNP are robust to the choice of estimation windows and arguably even higher. In contrast, the group of CYDLS, MSLS, MSSF and MSSO still fails to consistently improve out-of-sample portfolio performance. Instead, they behave similarly to first and second generation indices and mostly perform worse once longer estimation windows are considered. Concerning the portfolio turnover, the average trading decreases for longer estimation windows, which is plausible given the reduced relative impact of new information on parameter estimates. Finally, average commodity weights appear to be slightly higher, although this change is not consistent for each index or asset allocation strategy. To sum up, higher transaction costs and extended estimation windows worsen the average out-of-sample performance, except for some third generation indices, but generally, the results remain.

6 Conclusion

In recent years, the financial industry has constructed numerous indices that can provide commodity market exposure in a spectrum from purely passive investments to actively managed long–short strategies. Although these innovative indices are easily accessible for private and institutional investors and there exists evidence on the profitability of tactical asset allo-

[19] In case of 48- and 60-months estimation windows, the first out-of-sample return is calculated for February 2004 and February 2005, respectively, instead of February 2003.

cation in commodity markets, empirical research on enhanced indices is scarce. To this end, a comprehensive dataset of 21 commodity indices, that covers each generation and distinct strategies, is used to investigate whether investors can realize diversification benefits from adding first, second or third generation commodity indices to their portfolios.

We conduct various spanning tests and find that all indices of the first, most of the second and none of the third generation are spanned by stocks and bonds over the full sample. In contrast, all commodity indices reject the null of spanning when a sub-period of live-only index data are examined. These results suggest mixed diversification benefits of the first and second generation indices that are time-varying and more robust benefits of the third generation. Furthermore, the findings are largely unaffected by the inclusion of HML and SMB portfolios in the benchmark asset universe.

We then implement five different asset allocation strategies. In line with the evidence of mv spanning tests, first generation commodity indices do not offer portfolio diversification gains as indicated by almost identical out-of-sample SRs of benchmark and augmented portfolios. Average second generation indices can raise SRs for each strategy, except for $1/N$, yet the results are statistically not significant. Concerning third generation indices, the evidence is twofold: while half of the indices can increase the portfolio performance substantially, the other half does not provide any improvements at all. A plausible reason for the heterogenous findings is the inclusion of short futures contracts in third generation indices. In particular, they further emphasize the active management return component that is individual to each index and weaken the impact of the common exposure to spot price movements in commodities. Finally, qualitatively similar results are obtained when using higher transaction costs or different choices of the estimation window.

To sum up, the evidence presented in this paper is less clear-cut than the findings of Kremer (2015) and Daskalaki et al. (2017), who report increased portfolio performance of enhanced commodity indices that are more pronounced for the third generation. Although we indeed find that second and third generation indices are superior to first generation indices, the use of a larger and more diverse index sample reveals considerable differences within the group of third generation commodity indices.

7 Technical Appendix

7.1 Econometric Framework of Spanning Tests

The problem of spanning tests is usually presented within a regression framework, where the K benchmark asset returns are used to explain the returns of L test assets. This can be formalized as:

$$R_{2t} = \alpha + \beta R_{1t} + \varepsilon_t \qquad t = 1, 2, \ldots, T, \tag{3}$$

where $R_t = [R'_{1t},\, R'_{2t}]'$ represents the returns of the $L+K$ assets at time t, R_{1t} is a vector of the K benchmark asset returns, R_{2t} is an vector of the L test asset returns, ε_t is the vector of error terms, T is the length of the time series, α is the regression intercept and $\beta = (\beta_1, \ldots, \beta_K)$ are the regression slope parameters. In our study, only one commodity index is added to the benchmark assets and therefore $L = 1$.[20] Kan & Zhou (2012) note that R_t can be defined as either total returns or excess returns. However, Daskalaki & Skiadopoulos (2011) point out that, econometrically, excess returns are preferable in the regression framework when a risk-free asset complements the asset universe. Consequently, we use excess returns in the implementation of our mv spanning tests. Define $\delta = 1 - \beta 1_K$, where 1_K is a vector of ones. The null hypothesis of spanning in terms of restrictions on α and δ are:

$$H_0: \quad \alpha = 0, \quad \delta = 0. \tag{4}$$

If the spanning hypothesis holds, the inclusion of the test asset does not lead to more efficient portfolios than the initial asset universe. In turn, a rejection of the spanning hypothesis suggests diversification benefits from adding the test asset. Initially, Huberman & Kandel (1987) suggest testing for mv spanning with an LR test. Kan & Zhou (2012) take on this approach and expand it with a Wald test and LM test, which is also based on the regression in Eq. (3) and are outlined next. First, for simplicity, Eq. (3) can be rewritten in matrix notation as:

$$Y = XB + E, \tag{5}$$

[20] Kan & Zhou (2012) present a detailed description of the general case with arbitrary L test assets.

where Y is a $T \times 1$ vector of R_{2t}, X is a $T \times (K+1)$ matrix with its typical row as $[1,\ R'_{1t}]$, $B = [\alpha,\ \beta]'$ is a $(K+1) \times 1$ vector of regression coefficients, and E is a $T \times 1$ vector of ε_t. In order to estimate B and to derive the exact distributions of the Wald, LR and LM test statistics, Kan & Zhou (2012) point out that the following assumptions are required: α and β must be constant over time and conditional on $R_{1,t}$, the error terms ε_t have to be i.i.d. multivariate normal with mean zero and variance Φ. All three spanning tests can be estimated with maximum likelihood estimators of Eq. (5), which are given as follows:

$$\hat{B} \equiv [\hat{\alpha},\ \hat{\beta}]' = (X'X)^{-1}(X'Y), \tag{6}$$

$$\hat{\Phi} = \frac{1}{T}(Y - X\hat{B})'(Y - X\hat{B}). \tag{7}$$

Next, define $\Theta = [\alpha,\ \delta]'$. Since $\Theta = AB + C$, the null hypothesis of spanning can be reformulated as follows:

$$H_0: \quad \Theta = [\alpha,\ \delta]' = AB + C = 0_2, \tag{8}$$

where 0_2 is a 2×1 vector of zeros and A and C are given by:

$$A = \begin{bmatrix} 1 & 0'_K \\ 0 & -1'_K \end{bmatrix}, \qquad C = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

where 0_K is a $K \times 1$ vector of zeros. Using the unconstrained maximum likelihood estimators of Eq. (5), Θ can be estimated as $\hat{\Theta} \equiv [\hat{\alpha},\ \hat{\delta}]' = A\hat{B} + C$. Next, in order to derive the exact test statistics, Kan & Zhou (2012) show that the simplest way is to define two estimator matrices $\hat{G}$ and $\hat{H}$. The former is given by:

$$\hat{G} = TA(X'X)^{-1}A' = \begin{bmatrix} 1 + \hat{\mu}_1'\hat{V}_{11}^{-1}\hat{\mu}_1 & \hat{\mu}_1'\hat{V}_{11}^{-1}1_K \\ \hat{\mu}_1'\hat{V}_{11}^{-1}1_K & 1_K'\hat{V}_{11}^{-1}1_K \end{bmatrix}, \tag{9}$$

where $\hat{\mu}_1 = \frac{1}{T}\sum_{t=1}^{T} R_{1t}$ is a $K \times 1$ vector of the estimated expected returns of benchmark assets and $\hat{V}_{11} = \frac{1}{T}\sum_{t=1}^{T}(R_{1t} - \hat{\mu}_1)(R_{1t} - \hat{\mu}_1)'$ is the $K \times K$ estimated covariance matrix of benchmark asset returns. The second estimator matrix is given by:

$$\hat{H} = \hat{\Theta}\hat{\Phi}^{-1}\hat{\Theta}' = \begin{bmatrix} \hat{\alpha}'\hat{\Phi}^{-1}\hat{\alpha} & \hat{\alpha}'\hat{\Phi}^{-1}\hat{\delta} \\ \hat{\alpha}'\hat{\Phi}^{-1}\hat{\delta} & \hat{\delta}'\hat{\Phi}^{-1}\hat{\delta} \end{bmatrix}. \tag{10}$$

Kan & Zhou (2012) derive the Wald, LR and LM test statistics as a function of the two eigenvalues λ_1 and λ_2 of $\hat{H}\hat{G}^{-1}$, where $\lambda_1 \geq \lambda_2 \geq 0$. Accordingly, the test statistics are given by:

$$W = T(\lambda_1 + \lambda_2) \overset{a}{\sim} \chi_2^2, \tag{11}$$

$$LR = T\sum_{i=1}^{2} \ln(1+\lambda_i) \overset{a}{\sim} \chi_2^2, \tag{12}$$

$$LM = T\sum_{i=1}^{2} \frac{\lambda_i}{1+\lambda_i} \overset{a}{\sim} \chi_2^2. \tag{13}$$

Note that the three test statistics asymptotically follow a chi-squared distribution with degrees of freedom equal to the number of restrictions under the null hypothesis, hence two. It follows that the Wald, LR and LM tests are asymptotically equivalent regarding their conclusions about H_0. In finite samples, however, Berndt & Savin (1977) and Breusch (1979) show that $W \geq LR \geq LM$ is true, i.e. conflicting results are possible with the Wald test favoring rejection relative to the LR and LM tests. In order to circumvent this issue, it is sensible to implement all three tests.

7.2 Extensions of Basic Spanning Tests

The first extension suggested by Kan & Zhou (2012) is a step-down approach to disentangle return enhancement and risk reduction. In more detail, the authors propose first testing $\alpha = 0$ (F_1 test) and subsequently, conditional on $\alpha = 0$, testing $\delta = 0$ (F_2 test). In order to derive the respective test statistics, first define the following constants: $\hat{a} = \hat{\mu}'\hat{V}^{-1}\hat{\mu}$, $\hat{b} = \hat{\mu}'\hat{V}^{-1}1_K$, $\hat{c} = 1_K'\hat{V}^{-1}1_K$ and $\hat{d} = \hat{a}\hat{c} - \hat{b}^2$, where $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t$ and $\hat{V} = \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})'$. Likewise, the constants $\hat{a}_1$, $\hat{b}_1$, $\hat{c}_1$ and $\hat{d}_1$ are calculated for the K benchmark asset returns, using R_{1t}' instead of R_t. The F_1 test statistic is given by:

$$F_1 = (T-K-1)\left(\frac{|\bar{\Phi}|}{|\hat{\Phi}|} - 1\right) = (T-K-1)\left(\frac{\hat{a} - \hat{a}_1}{1+\hat{a}_1}\right), \tag{14}$$

where $|\bar{\Phi}|$ is the constraint estimate of Φ under the $\alpha = 0$ constraint and $|\hat{\Phi}|$ is the unconstraint estimate of Φ. Under the null hypothesis, F_1 follows a central F-distribution with 1 and $(T-K-1)$ degrees of freedom. Rejecting the H_0 of the F_1 test implies a return enhancement of the optimal portfolio in the augmented asset universe compared to the initial one. Next, the F_2 test statistic is defined as:

$$F_2 = (T-K)\left(\frac{|\tilde{\Phi}|}{|\bar{\Phi}|} - 1\right) = (T-K)\left[\left(\frac{\hat{c}+\hat{d}}{\hat{c}_1+\hat{d}_1}\right)\left(\frac{1+\hat{a}_1}{1+\hat{a}}\right) - 1\right], \tag{15}$$

where $\tilde{\Phi}$ is the constrained estimate of Φ under both constraints, $\alpha = 0$ and $\delta = 0$. Under the null hypothesis, F_2 follows a central F-distribution with 1 and $(T-K)$ degrees of freedom. Rejecting the H_0 of Eq. (15) implies a risk reduction of the optimal portfolio in the augmented asset universe compared to the initial one. As indicated above, the F_1 and F_2 tests allow a more precise breakdown of the actual source of diversification benefits, i.e. whether test assets improve returns or reduce risks. Therefore, the step-down test provides additional information complementary to the basic spanning tests.

The second issue addressed is the heteroscedasticity of the error term which biases the maximum likelihood-based tests. Assuming R_t to have finite fourth moments and defining $x_t = [1,\ R'_{1t}]'$ as well as $\varepsilon_t = R_{2t} - B'x_t$, the moment conditions of the GMM estimation for B are given by:

$$E[g_t] = E[x_t \otimes \varepsilon_t] = 0_{(K+1)}, \tag{16}$$

where $\otimes$ denotes the Kronecker product of two matrices. The sample moments are given by:

$$\bar{g}_T(B) = \frac{1}{T}\sum_{t=1}^{T} x_t \otimes (R_{2t} - B'x_t). \tag{17}$$

Therefore, B can be estimated using the GMM by minimizing the objective function $J_T = \bar{g}_T(B)'S_T^{-1}\bar{g}_T(B)$, where S_T is a consistent estimate of the weighting matrix $S_0 = E[g_t g'_t]$ under the assumption that g_t is serially uncorrelated. However, Kan & Zhou (2008) point out that in this case the GMM system is exactly identified, i.e. the amount of moment conditions equals the number of parameters to estimate, and therefore the procedure can be simplified. In fact, it follows that the unconstrained estimates of $\hat{B}$ and $\hat{\Theta}$ are independent of S_T and thus remain the same as derived earlier. Newey & West (1987) show that for the special case of a GMM estimation with a linear model and linear constraints, the Wald, LR and LM tests are identical. Therefore, it is sufficient to calculate the Wald test under the GMM, which does not require a constrained estimate of B.

The GMM-Wald test statistic is given by:

$$W_a = T\mathrm{vec}(\hat{\Theta}')'(A_T S_T A_T')^{-1}\mathrm{vec}(\hat{\Theta}') \overset{a}{\sim} \chi^2_2, \tag{18}$$

where *vec* denotes a matrix vectorization, S_T is calculated using the unconstrained estimate of B as $S_T = E\left[(x_t \otimes \hat{\varepsilon}_t)(x_t \otimes \hat{\varepsilon}_t)'\right]$ and A_T is given by:

$$A_T = \begin{bmatrix} 1+\hat{a}_1 & -\hat{\mu}_1 \hat{V}_{11}^{-1} \\ \hat{b}_1 & -1_K' \hat{V}_{11}^{-1} \end{bmatrix}. \tag{19}$$

Note that the GMM-Wald statistics in Eq. (18) can be interpreted similarly as the spanning tests outlined in Eq. (11)–(13), i.e. a rejection of the null signals a statistically significant shift in the augmented efficient frontier.

7.3 Details on Portfolio Strategies

First, portfolios consistent with the classical Markowitz (1952) framework are constructed, i.e. the investor only cares about the mean and variance of portfolio returns and optimizes their tradeoff. The well known mv optimization problem is given by:

$$\max_{\omega} U = \omega'\mu - \frac{\gamma}{2}\omega'\Sigma\omega, \tag{20}$$

where U, ω, Σ and γ denote the investor's utility, the vector of portfolio weights, the covariance matrix of returns and the investor's risk aversion, respectively. As is well known, the unrestricted mv optimization is highly sensitive to errors in parameter estimates, which can result in extreme portfolio allocations. For this reason, unconstrained portfolios derived from Eq. (20) are likely not implemented by practitioners and therefore portfolio constraints are introduced. Jagannathan & Ma (2003) show that portfolio constraints alleviate the issue of extreme allocations and can also enhance performance. In a first step, the usual budget and short sell restrictions are implemented for each asset allocation strategy covered. These constraints can be formalized as $\sum_{i=1}^{N} \omega_i = 1$ and $0 \leq \omega_i \leq 1$, respectively, where ω_i is an element of the vector ω and N is the total amount of assets considered. In addition, the mv optimization problem of Eq. (20) is also restricted by an upper volatility bound which is given by:

$$\sqrt{\omega' \Sigma \omega} \leq \hat{\sigma}_C, \tag{21}$$

where $\hat{\sigma}_C$ denotes the portfolio maximum volatility. This specific restriction of the mv strategy enhances the comparability with other strategies and is presumably close to realistic asset management. Furthermore, to determine optimal mv portfolio weights according to Eq. (18), an assumption about the investor's risk aversion is necessary. To this end, we follow Bessler & Wolff (2015) and consider an aggressive ($\gamma = 2$, $\hat{\sigma}_C = 5\%$) and a conservative ($\gamma = 10$, $\hat{\sigma}_C = 15\%$) investor.[21] Finally, both mv strategies are implemented out-of-sample by estimating μ and Σ with their sample counterparts, $\hat{\mu}$ and $\hat{\Sigma}$, and maximizing the investor's utility with respect to ω subject to the constraints. Additionally, the mv strategies are applied in-sample to obtain the maximum achievable diversification benefits of commodities as an upper limit benchmark. The in-sample mv implementation is similar to the out-of-sample case, except that asset return estimates for the next month are replaced by the respective realized returns of that month.

The second out-of-sample asset allocation strategy is $1/N$ with monthly rebalancing, i.e. an equally weighted investment over all available N assets. The portfolio weights of the $1/N$ strategy are given by $\omega_i^{1/N} = 1/N$. This strategy is frequently applied in practice, predominantly by private investors (Benartzi & Thaler, 2001), and its simplicity does not necessarily makes it inferior to more complex investment approaches (DeMiguel et al., 2009). Practically, the implementation of $1/N$ is simple because it does not require any parameter estimates and, as a result, rules out estimation errors.

Next, risk parity (RP) is considered as a third asset allocation strategy that aims at equalizing each asset's contribution to the overall portfolio risk (Asness, Frazzini, & Pedersen, 2012). Accordingly, a RP strategy does not diversify by relative investments in different asset classes but by the risk an asset class adds to the whole portfolio. As discussed by Anderson, Bianchi, & Goldberg (2012), RP strategies became popular after the financial crisis in 2008 and have

[21] Similar parameter choices are also used in Daskalaki & Skiadopoulos (2011).

ever since been applied by long-term investors, such as pension and endowment funds. From an implementation point of view, the simple RP strategy is applied as in Bessler & Wolff (2015). It does not consider asset return correlations and weights each asset anti-proportional to its sample variance. The simple RP portfolio weights are simply calculated as:

$$\omega_i^{RP} = \frac{1/\hat{\sigma}_i^2}{\sum_{i=1}^{N}(1/\hat{\sigma}_i^2)}. \tag{22}$$

Finally, the RRT strategy introduced by Kirby & Ostdiek (2012) represents the fourth asset allocation strategy we consider. Under this approach, portfolio weights are determined by the reward-to-risk ratio of the respective assets. This ratio is calculated as the sample mean of an asset divided by its sample variance as follows:

$$\omega_i^{RRT} = \frac{\hat{\mu}_i^+/\hat{\sigma}_i^2}{\sum_{i=1}^{N}(\hat{\mu}_i^+/\hat{\sigma}_i^2)}, \tag{23}$$

where $\hat{\mu}_i^+ = \max(\hat{\mu}_i, 0)$. By defining $\hat{\mu}_i^+$ and neglecting the off-diagonal elements of $\hat{\Sigma}$, the RRT strategy aims to mitigate extreme allocations, does not allow for short sales and rules out investments in assets with negative expected returns. In the case that all estimated asset returns are negative, i.e. no weights can be determined by Eq. (23), wealth is equally split among all the available assets.

Bibliography

Anderson, R. M., Bianchi, S. W., & Goldberg, L. R. (2012). Will my risk parity strategy outperform? *Financial Analysts Journal*, *68*(6), 75–93.

Anson, M. J. (1999). Maximizing utility with commodity futures diversification. *Journal of Portfolio Management*, *25*(4), 86–94.

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2012). Leverage aversion and risk parity. *Financial Analysts Journal*, *68*(1), 47–59.

Barclays (2014). Barclays backwardation long/short TR index. https://www.sec.gov/Archives/edgar/data/312070/000119312514082986/d686424dfwp.htm.

Belousova, J., & Dorfleitner, G. (2012). On the diversification benefits of commodities from the perspective of euro investors. *Journal of Banking & Finance*, *36*(9), 2455–2472.

Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review*, *91*(1), 79–98.

Berndt, E. R., & Savin, N. E. (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 & Finance*, *60*, 1–20.

Bloomberg (2017). Index methodology. The bloomberg commodity index family. https://www.bloomberg.com/.

Bloomberg Barclays Index (2017). US aggregate index. https://www.bloomberg.com/.

Bodie, Z., & Rosansky, V. I. (1980). Risk and return in commodity futures. *Financial Analysts Journal*, *36*(3), 27–39.

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

Büyükşahin, B., & Robe, M. A. (2014). Speculators, commodities and cross-market linkages. *Journal of International Money and Finance*, *42*, 38–70.

Credit Suisse (2010). Credit suisse MASTRO index. „Multi-algo strategy targeted risk overlay“. https://research-doc.credit-suisse.com/.

CYD Research (2013). CYD commodity index guide. https://www.vontobel.com/.

Daskalaki, C., & Skiadopoulos, G. (2011). Should investors include commodities in their portfolios after all? New evidence. *Journal of Banking & Finance*, *35*(10), 2606–2626.

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

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.

Deutsche Bank (2008). DBIQ index guide. DBLCI commodity indices. https://index.db.com/dbiqweb2/home.do?redirect=homepage.

Domanski, D., & Heath, A. (2007). Financial investors and commodity markets. *BIS Quarterly Review*, *March*, 53–67.

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

Erb, C. B., & Harvey, C. R. (2006). The strategic and tactical value of commodity futures. *Financial Analysts Journal*, *62*(2), 69–97.

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

Geman, H. (2005). *Commodities and commodity derivatives: Modeling and pricing for agriculturals, metals and energy*. Chichester: John Wiley & Sons.

Gorton, G., & Rouwenhorst, K. G. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal*, *62*(2), 47–68.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029–1054.

Huang, J.-z., & Zhong, Z. K. (2013). Time variation in diversification benefits of commodity, REITs, and TIPS. *Journal of Real Estate Finance and Economics*, *46*, 1–41.

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

Irwin, S. H., & Sanders, D. R. (2011). Index funds, financialization, and commodity futures markets. *Applied Economic Perspectives and Policy*, *33*(1), 1–31.

Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, *58*(4), 1651–1683.

Jensen, G. R., Johnson, R. R., & Mercer, J. M. (2000). Efficient use of commodity futures in diversified portfolios. *Journal of Futures Markets*, *20*(5), 489–506.

Jensen, G. R., Johnson, R. R., & Mercer, J. M. (2002). Tactical asset allocation and commodity futures. *Journal of Portfolio Management*, *28*(4), 100–111.

Jensen, G. R., & Mercer, J. M. (2011). Commodities as an investment. *The Research Foundation of CFA Institute Literature Review*, *6*(2), 1–33.

Jobson, J. D., & Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance*, *36*(4), 889–908.

Kan, R., & Zhou, G. (2008). Tests of mean-variance spanning. OLIN Working Paper 99-05.

Kan, R., & Zhou, G. (2012). Tests of mean-variance spanning. *Annals of Economics & Finance*, *13*(1).

Kirby, C., & Ostdiek, B. (2012). It's all in the timing: Simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis*, *47*(2), 437–467.

Kremer, P. J. (2015). Comparing three generations of commodity indices: New evidence for portfolio diversification. *Alternative Investment Analyst Review*, *3*(4), 30–43.

Lombardi, M., & Ravazzolo, F. (2016). On the correlation between commodity and equity returns: Implications for portfolio allocation. *Journal of Commodity Markets*, *2*, 45–57.

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

Memmel, C. (2003). Performance hypothesis testing with the Sharpe ratio. *Finance Letters*, *1*(1), 21–23.

Merrill Lynch (2006). Selecting a commodity index. Global Commodity Paper 4.

Merton, R. C. (1972). An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis*, *7*(4), 1851–1872.

Miffre, J. (2012). Comparing first, second and third generation commodity indices. Working Paper.

Miffre, J. (2016). Long-short commodity investing: A review of the literature. *Journal of Commodity Markets*, *1*, 3–13.

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

Morningstar (2013). Construction rules for the Morningstar commodity index family. http://advisor.morningstar.com/Principia/pdf/MorningstarCommodityMethodology.pdf.

Newey, W. K., & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, *28*(3), 777–787.

Ohashi, K., & Okimoto, T. (2016). Increasing trends in the excess comovement of commodity prices. *Journal of Commodity Markets*, *1*, 48–64.

Paschke, R., Prokopczuk, M., & Wese Simen, C. (2017). Curve momentum. Working Paper.

Rallis, G., Miffre, J., & Fuertes, A.-M. (2013). Strategic and tactical roles of enhanced commodity indices. *Journal of Futures Markets*, *33*(10), 965–992.

Scherer, B., & He, L. (2008). The diversification benefits of commodity futures indexes: A mean-variance spanning test. In F. J. Fabozzi, R. Fuess, & D. G. Kaiser (Eds.) *The Handbook of Commodity Investing*, chap. 10, (pp. 241–265). Hoboken, NJ: John Wiley & Sons.

Silvennoinen, A., & Thorp, S. (2013). Financialization, crisis and commodity correlation dynamics. *Journal of International Financial Markets, Institutions and Money*, *24(C)*, 42–65.

S&P Dow Jones Indices (2017). S&P GSCI methodology.

Stoll, H. R., & Whaley, R. E. (2011). Commodity index investing: Speculation or diversification? *Journal of Alternative Investments*, *14*(1), 50–60.

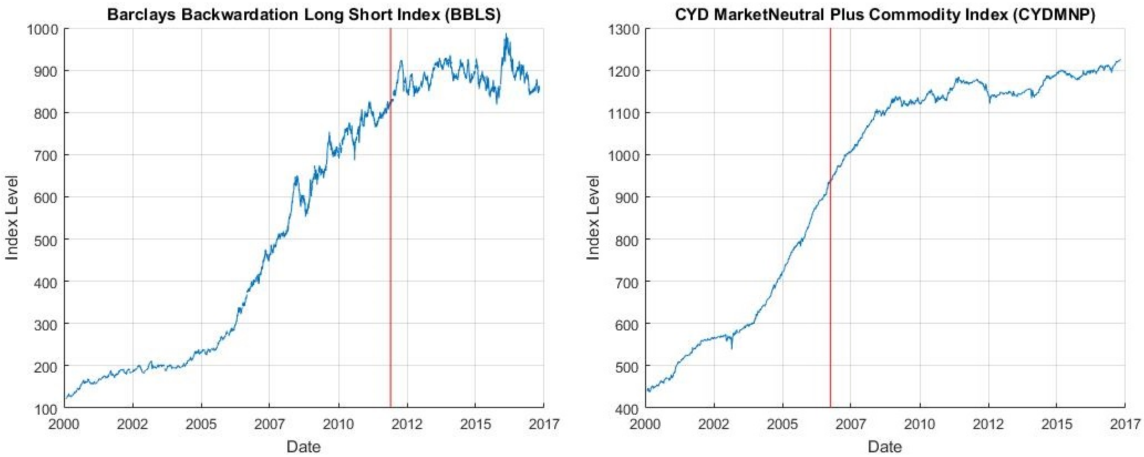
SummerHaven Index Management (n.d.). SummerHaven dynamic commodity index („SDCI“). Index methodology. https://summerhavenindex.com/sdci/.

Szymanowska, M., DeRoon, F., Nijman, T., & Goorbergh, R. (2014). An anatomy of commodity futures risk premia. *Journal of Finance*, *69*(1), 453–482.

Tang, K., & Xiong, W. (2012). Index investment and the financialization of commodities. *Financial Analysts Journal*, *68*(5), 54–74.

Vescore Indices (2016). CYD diversified longshort commodity index. Leitfaden und Informationen bzgl. des Indexes. https://www.vontobel.com/en-de.

Figure 1: Plots of Third Generation Commodity Indices



Notes: This figure illustrates the daily time series of two third generation commodity indices, the BBLS and the CYDMNP, for the February 2000–April 2017 period. The red bars highlight the launch dates of both indices. Source: Own representation.

Table 1: Descriptive Statistics of Assets (February 2000 – April 2017)

Asset	Annualized Mean (%)	t-stat	Annualized Volatility (%)	Sharpe Ratio	Skewness	Kurtosis	JB (p-value)	Engle (p-value)
First Generation Commodity Indices								
BCOM	2.20	(0.55)	16.52	0.04	-0.48	4.85	0.001	0.019
DBLCI	3.93	(0.79)	20.75	0.11	-0.42	4.18	0.004	0.004
S&P-GSCI	1.09	(0.20)	23.05	-0.02	-0.39	4.14	0.005	0.004
Second Generation Commodity Indices								
CYDLO	7.78	(2.11)	15.31	0.41	-0.30	3.94	0.015	0.000
DBLCI-MR	5.61	(1.19)	19.62	0.21	-0.46	3.94	0.007	0.007
DBLCI-OY	6.27	(1.39)	18.76	0.25	-0.53	4.61	0.001	0.006
DBLCI-OYBA	7.51	(1.92)	16.23	0.37	-0.68	5.89	0.001	0.039
DBLCI-OYBR	8.08	(1.82)	18.44	0.35	-0.57	5.02	0.001	0.032
MLCX	6.43	(1.24)	21.55	0.23	-0.39	4.28	0.003	0.020
MSLF	6.29	(2.26)	11.57	0.41	0.19	5.06	0.001	0.006
MSLO	5.19	(1.20)	17.94	0.20	-0.39	4.50	0.002	0.025
SDCI	10.51	(2.82)	15.49	0.58	-0.70	6.56	0.001	0.009
Third Generation Commodity Indices								
BBLS	11.93	(4.61)	10.75	0.97	0.23	2.98	0.362	0.301
CSMOVERS	9.69	(2.73)	14.77	0.55	0.47	5.50	0.001	0.646
CSMOVERSMN	7.28	(2.93)	10.30	0.56	0.22	3.67	0.063	0.816
CYDDLS	14.65	(5.77)	10.54	1.25	0.27	3.36	0.139	0.983
CYDLS	4.04	(1.96)	8.58	0.29	0.37	3.26	0.066	0.375
CYDMNP	5.96	(6.99)	3.55	1.29	0.35	12.16	0.001	0.000
MSLS	5.81	(2.07)	11.66	0.36	0.38	4.81	0.001	0.021
MSSF	1.12	(0.80)	5.82	-0.08	0.71	8.37	0.001	0.000
MSSO	-1.53	(-0.36)	17.43	-0.18	0.59	4.40	0.001	0.018
Averages								
First Generation	2.41	-	20.11	0.04	-0.43	4.39	-	-
Second Generation	7.07	-	17.21	0.33	-0.42	4.87	-	-
Third Generation	6.55	-	10.38	0.56	0.40	5.39	-	-
Benchmarks								
Barclays Bonds	5.18	(6.23)	3.46	1.05	-0.38	4.38	0.003	0.004
S&P 500	6.15	(1.73)	14.74	0.31	-0.56	4.16	0.002	0.000
HML	4.31	(1.59)	11.24	0.24	0.19	5.83	0.001	0.000
SMB	3.48	(1.23)	11.70	0.16	0.96	14.63	0.001	0.000
T-Bill	1.59	(11.78)	0.56	-	1.04	2.71	0.001	0.000

Notes: This table reports descriptive statistics for first, second and third generation commodity indices, equity and bond indices, value and size portfolios and the risk-free rate over the period from February 2000 to April 2017. Entries are based on monthly total return data. t-stat reports the t-statistic of a standard t-test with the null of an annual mean equal to zero. The Sharpe ratio is calculated as annualized time series mean of excess returns divided by the annualized volatility of net excess returns. Jarque-Bera (JB) p-values are reported to test for normality of returns. The null states that asset returns follow a normal distribution. Engle p-values are reported to test for conditional heteroscedasticity in the residuals. The null states that residuals show no conditional heteroscedasticity.

Table 2: Correlation Matrix of Asset Returns (February 2000 – April 2017)

Asset	BCOM	DBLCI	S&P-GSCI	CYDLO	DBLCI-MR	DBLCI-OY	DBLCI-OYBA	DBLCI-OYBR	MLCX	MSLF	MSLO	SDCI	BBLS	CSMOVERS	CSMOVERSMN	CYDDLS	CYDLS	CYDMNP	MSLS	MSSF	MSSO	Barclays Bonds	S&P 500	HML	SMB	T-Bill
BCOM	1.00																									
DBLCI	0.90**	1.00																								
S&P-GSCI	0.90*	0.96**	1.00																							
CYDLO	0.88**	0.85**	0.87**	1.00																						
DBLCI-MR	0.82**	0.87**	0.81**	0.70**	1.00																					
DBLCI-OY	0.91**	0.97**	0.93**	0.86**	0.86**	1.00																				
DBLCI-OYBA	0.95**	0.91**	0.87**	0.89**	0.82**	0.95**	1.00																			
DBLCI-OYBR	0.93**	0.96**	0.95**	0.90**	0.84**	0.98**	0.97**	1.00																		
MLCX	0.92**	0.96**	0.98**	0.88**	0.83**	0.95**	0.91**	0.97**	1.00																	
MSLF	0.78**	0.75**	0.74**	0.78**	0.60**	0.73**	0.75**	0.76**	0.75**	1.00																
MSLO	0.97**	0.93**	0.93**	0.91**	0.83**	0.93**	0.94**	0.96**	0.96**	0.80**	1.00															
SDCI	0.89**	0.81**	0.79**	0.88**	0.72**	0.84**	0.92**	0.88**	0.82**	0.77**	0.89**	1.00														
BBLS	0.08	0.18**	0.17*	0.26**	-0.01	0.15*	0.14*	0.17*	0.16*	0.32**	0.09	0.25**	1.00													
CSMOVERS	0.06	0.08	0.07	0.11	0.02	0.06	0.07	0.07	0.07	0.43**	0.08	0.15*	0.42**	1.00												
CSMOVERSMN	0.21**	0.25**	0.24**	0.24**	0.19**	0.23**	0.22**	0.24**	0.26**	0.45**	0.24**	0.29**	0.57**	0.77**	1.00											
CYDDLS	0.01	0.11	0.16*	0.22**	0.01	0.12	0.07	0.13	0.15*	0.10	0.07	0.14	0.61**	0.15*	0.34**	1.00										
CYDLS	0.00	0.09	0.13	0.32**	-0.11	0.06	0.00	0.07	0.10	0.31**	0.05	0.06	0.51**	0.29**	0.27**	0.42**	1.00									
CYDMNP	-0.21**	-0.23**	-0.31**	-0.17*	-0.14*	-0.11	-0.10	-0.14*	-0.23**	-0.19**	-0.21**	-0.08	0.05	0.01	-0.02	0.06	-0.09	1.00								
MSLS	0.49**	0.50**	0.53**	0.53**	0.34**	0.47**	0.45**	0.50**	0.52**	0.88**	0.50**	0.48**	0.40**	0.58**	0.50**	0.15*	0.45**	-0.15*	1.00							
MSSF	-0.58**	-0.48**	-0.42**	-0.50**	-0.51**	-0.52**	-0.59**	-0.52**	-0.46**	-0.23**	-0.57**	-0.56**	0.18*	0.32**	0.13	0.11	0.28**	0.10	0.26**	1.00						
MSSO	-0.94**	-0.88**	-0.89**	-0.87**	-0.81**	-0.89**	-0.91**	-0.92**	-0.91**	-0.71**	-0.96**	-0.84**	0.07	0.04	-0.09	0.02	0.02	0.24**	-0.40**	0.63**	1.00					
Barclays Bonds	0.05	0.01	-0.03	0.03	-0.03	0.01	0.06	0.00	-0.02	0.05	0.04	0.07	0.08	-0.03	-0.02	-0.03	0.07	0.03	-0.02	-0.14*	-0.02	1.00				
S&P 500	0.36**	0.30**	0.29**	0.28**	0.35**	0.35**	0.37**	0.35**	0.32**	0.11	0.32**	0.36**	-0.01	-0.17*	-0.11	-0.02	-0.24**	0.01	-0.09	-0.41**	-0.32**	-0.10	1.00			
HML	0.10	0.03	0.08	0.02	0.10	0.04	0.06	0.07	0.09	-0.02	0.08	0.05	-0.03	-0.01	-0.05	-0.07	-0.15*	0.12	-0.03	0.02	-0.04	-0.04	0.01	1.00		
SMB	0.07	0.16*	0.11	0.14	0.13	0.15*	0.11	0.12	0.12	0.05	0.06	0.05	0.13	0.02	0.06	0.21**	0.16*	-0.07	0.07	0.03	-0.06	-0.07	0.13	-0.28**	1.00	
T-Bill	0.09	0.12	0.08	0.13	0.12	0.13	0.12	0.12	0.11	0.13	0.09	0.17*	0.25**	0.19**	0.27**	0.15*	0.17*	0.34**	0.15*	0.14*	0.01	0.12	-0.10	0.18**	-0.04	1.00

Notes: This table reports the Pearson's correlation coefficients for first, second and third generation commodity, equity and bond indices, value and size portfolios and the risk-free rate over the period from February 2000 to April 2017. * and ** indicate correlation values significantly different from 0 at the 5% and 1% level, respectively.

Table 3: Results of Spanning Tests for Commodity Indices (February 2000 – April 2017)

Commodity	Wald	LR	LM	GMM-Wald	F_1	F_2
First Generation Commodity Indices						
BCOM	1.530	1.525	1.519	1.562	0.381	1.131
	(0.465)	(0.467)	(0.468)	(0.458)	(0.538)	(0.289)
DBLCI	1.439	1.434	1.429	0.924	0.000	1.425
	(0.487)	(0.488)	(0.489)	(0.630)	(0.991)	(0.234)
S&P-GSCI	3.052	3.030	3.008	2.204	0.126	2.894*
	(0.217)	(0.220)	(0.222)	(0.332)	(0.723)	(0.090)
Second Generation Commodity Indices						
CYDLO	3.703	3.670	3.638	2.320	1.357	2.288
	(0.157)	(0.160)	(0.162)	(0.313)	(0.245)	(0.132)
DBLCI-MR	3.071	3.048	3.026	2.276	0.271	2.765*
	(0.215)	(0.218)	(0.220)	(0.321)	(0.603)	(0.098)
DBLCI-OY	1.727	1.719	1.712	1.067	0.291	1.415
	(0.422)	(0.423)	(0.425)	(0.586)	(0.590)	(0.236)
DBLCI-OYBA	0.972	0.970	0.967	0.577	0.569	0.390
	(0.615)	(0.616)	(0.616)	(0.749)	(0.452)	(0.533)
DBLCI-OYBR	2.273	2.261	2.248	1.286	0.963	1.277
	(0.321)	(0.323)	(0.325)	(0.526)	(0.328)	(0.260)
MLCX	2.193	2.181	2.170	1.299	0.356	1.811
	(0.334)	(0.336)	(0.338)	(0.522)	(0.552)	(0.180)
MSLF	10.755***	10.484***	10.223***	16.408***	1.765	8.801***
	(0.005)	(0.005)	(0.006)	(0.000)	(0.185)	(0.003)
MSLO	1.012	1.009	1.007	0.743	0.046	0.956
	(0.603)	(0.604)	(0.604)	(0.690)	(0.831)	(0.329)
SDCI	3.110	3.087	3.064	2.151	2.779*	0.283
	(0.211)	(0.214)	(0.216)	(0.341)	(0.097)	(0.595)
Third Generation Commodity Indices						
BBLS	22.087***	20.986***	19.958***	19.854***	13.466***	7.825***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.006)
CSMOVERS	25.480***	24.030***	22.688***	11.797***	7.508***	17.062***
	(0.000)	(0.000)	(0.000)	(0.003)	(0.007)	(0.000)
CSMOVERSMN	37.702***	34.636***	31.893***	40.832***	7.135***	29.148***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.008)	(0.000)
CYDDLS	42.255***	38.452***	35.092***	32.401***	26.648***	13.327***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CYDLS	38.144***	35.009***	32.209***	23.527***	1.854	35.588***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.175)	(0.000)
CYDMNP	198.767***	139.323***	101.400***	228.457***	26.193***	151.120***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MSLS	26.683***	25.098***	23.636***	31.698***	3.183*	22.870***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.076)	(0.000)
MSSF	194.768***	137.273***	100.349***	127.884***	1.200	190.560***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.275)	(0.000)
MSSO	23.239***	22.024***	20.893***	18.988***	0.022	22.989***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.881)	(0.000)

Notes: This table reports test statistics and respective p-values (in parentheses) for several tests of the null hypotheses that stocks and bonds span commodities. Columns two to four present the Wald, Likelihood ratio (LR) and Lagrange multiplier (LM) tests. The fifth column reports the results of the Wald test based on the GMM estimation (GMM-Wald). Columns six and seven report the results of the step-down procedure: F_1 tests for a return enhancement and F_2 for a risk reduction. Stocks are represented by the S&P 500 Index, bonds by the Bloomberg Barclays US Aggregate Bond Index and commodities by the respective index presented in each row. The analysis is based on monthly excess returns over the 1-month U.S. Treasury Bill and covers the period from February 2000 to April 2017. *, ** and *** indicate significant entries at the 10%, 5% and 1% significance level, respectively.

Table 4: In-Sample Portfolio Optimization with Commodities

	Benchmark	Benchmark portfolio complemented with commodity index																				
		BCOM	DBLCI	S&P-GSCI	CYDLO	DBLCI-MR	DBLCI-OY	DBLCI-OYBA	DBLCI-OYBR	MLCX	MSLF	MSLO	SDCI	BBLS	CSMOVERS	CSMOVERSMN	CYDDLS	CYDLS	CYDMNP	MSLS	MSSF	MSSO
Panel A: mv (aggressive)																						
Return (%)	31.99	42.52	45.95	45.42	45.10	48.23	47.51	45.43	48.22	46.90	39.87	45.13	45.19	42.61	48.62	41.77	44.40	38.28	33.84	41.36	35.18	50.31
Volatility (%)	6.99	8.26	8.50	8.38	8.31	9.23	8.82	8.50	8.77	8.50	7.83	8.48	8.71	7.32	9.25	7.26	7.40	6.57	6.62	7.88	6.71	9.02
Sharpe ratio	4.32	4.95*	5.23**	5.23**	5.24**	5.08	5.22**	5.17**	5.33**	5.32**	4.87*	5.13*	5.02*	5.61***	5.07	5.52***	5.75***	5.53***	4.84***	5.02*	4.94***	5.35*
Port. turnover (%)	146.99	153.59	146.84	147.76	147.25	149.62	146.18	144.49	142.15	145.90	151.25	145.48	146.72	157.09	143.37	151.23	152.40	154.44	153.03	153.23	156.35	151.20
Avg. com. weight (%)	-	23.49	22.33	20.03	25.93	28.12	26.49	28.54	28.76	24.26	22.22	26.10	26.23	29.36	30.36	26.78	33.35	20.51	12.73	21.14	13.57	25.34
Panel B: mv (conservative)																						
Return (%)	18.64	21.43	21.86	21.64	22.32	22.70	22.36	22.48	22.53	21.95	22.12	21.99	22.88	24.13	24.93	24.22	25.10	22.94	20.79	23.01	21.99	24.42
Volatility (%)	4.22	4.31	4.28	4.24	4.27	4.48	4.30	4.41	4.29	4.23	4.37	4.30	4.54	4.47	4.54	4.45	4.44	4.01	3.93	4.24	4.37	4.23
Sharpe ratio	4.09	4.67***	4.82***	4.81***	4.94***	4.82***	4.91***	4.83***	4.97***	4.89***	4.77***	4.82***	4.78***	5.13***	5.19***	5.17***	5.36***	5.35***	4.96***	5.09***	4.67*	5.42***
Port. turnover (%)	98.84	102.19	98.77	99.32	101.36	101.08	98.57	100.16	98.34	99.03	106.82	100.26	102.63	109.74	104.05	112.91	107.90	118.36	127.82	109.39	129.21	102.59
Avg. com. weight (%)	-	7.00	5.88	5.03	8.65	7.76	7.23	9.37	7.71	6.10	11.89	7.44	9.78	17.08	12.34	16.46	18.29	18.80	27.24	12.45	20.79	10.15

Notes: This table reports the results of in-sample optimized mean–variance portfolios for a benchmark (stock–bond) portfolio and the benchmark portfolio complemented with one out of 21 commodity indices at a time. The results are derived from a rolling sample approach with a 36-month estimation window and cover the full sample from February 2000 to April 2017. Return and volatility denote the annualized time series mean and volatility of monthly net returns, respectively. The Sharpe ratio is calculated as annualized time series mean of net excess returns divided by the annualized time series volatility of net excess returns. 'Port. turnover' denotes the portfolio turnover, calculated as the average change in weights over all rebalancing points. 'Avg. com. weight' denotes the average weight in the commodity index over the full sample. Panel A describes the results of an aggressive mv strategy with a risk aversion coefficient of 2 and an upper volatility bound of 15%. Panel B describes the results of a conservative mv strategy with a risk aversion coefficient of 10 and an upper volatility bound of 5%. Italic font indicates an absolute increase in the Sharpe ratio of the augmented portfolio relative to the benchmark case. *, ** and *** indicate a significant increase in the Sharpe ratio of the augmented portfolio relative to the benchmark case at the 10%, 5% and 1% level, respectively.

Table 5: Out-of-Sample Portfolio Optimization with Commodities

		Benchmark portfolio complemented with commodity index																				
	Benchmark	BCOM	DBLCI	S&P-GSCI	CYDLO	DBLCI-MR	DBLCI-OY	DBLCI-OYBA	DBLCI-OYBR	MLCX	MSLF	MSLO	SDCI	BBLS	CSMOVERS	CSMOVERSMN	CYDDLS	CYDLS	CYDMNP	MSLS	MSSF	MSSO
Panel A: mv (aggressive)																						
Return (%)	7.03	8.60	9.70	8.26	12.19	12.25	13.14	12.26	13.13	10.96	7.65	9.41	14.58	13.30	12.43	10.66	16.68	6.59	8.76	5.74	6.19	4.70
Volatility (%)	8.50	11.16	12.73	10.70	12.35	13.18	13.39	12.71	13.31	12.20	10.68	12.25	12.97	9.88	12.42	9.49	9.96	8.45	7.04	9.89	8.47	8.76
Sharpe ratio	0.68	0.66	0.66	0.65	_0.88_	_0.84_	_0.89_	_0.87_	_0.89_	_0.79_	0.60	0.66	_1.03_	_1.22**_	_0.90_	_0.99_	_1.54***_	0.63	_1.06**_	0.46	0.58	0.40
Port. turnover (%)	14.11	14.79	12.32	12.62	8.47	12.57	8.67	9.77	9.47	11.15	20.46	13.32	11.42	13.80	13.83	13.26	13.47	20.71	16.43	22.78	17.69	20.69
Avg. com. weight (%)	-	22.12	26.53	14.38	32.64	34.71	32.97	33.49	31.84	25.69	28.76	28.13	41.03	42.28	40.49	36.20	68.59	16.02	29.83	22.35	2.03	9.64
Panel B: mv (conservative)																						
Return (%)	4.90	5.58	5.89	5.44	6.71	6.33	6.75	6.92	6.89	6.27	5.62	6.03	7.64	7.41	6.92	7.03	9.68	4.70	7.03	4.61	4.27	5.01
Volatility (%)	4.59	4.84	5.32	4.78	5.25	5.07	5.38	5.38	5.40	5.06	4.96	5.12	5.50	4.88	5.17	4.91	5.34	4.71	3.60	4.74	4.45	4.72
Sharpe ratio	0.79	_0.89_	_0.88_	_0.88_	_1.04_	_1.01_	_1.03_	_1.06_	_1.05_	_0.99_	_0.88_	_0.93_	_1.17*_	_1.28**_	_1.10_	_1.19*_	_1.58***_	0.73	_1.64***_	0.71	0.68	_0.79_
Port. turnover (%)	12.36	11.15	10.80	11.22	10.18	11.57	10.14	10.33	10.16	9.99	13.25	10.87	11.03	14.74	12.36	12.64	8.96	16.16	10.52	14.60	15.41	14.52
Avg. com. weight (%)	-	6.81	7.03	4.04	9.96	9.80	9.40	10.99	9.22	6.81	12.29	8.37	14.21	20.39	14.81	20.66	36.38	11.14	35.56	10.54	4.80	5.57
Panel C: 1/N with rebalancing																						
Return (%)	4.54	3.65	3.95	3.30	4.81	4.65	4.56	5.02	4.98	4.29	4.49	4.24	5.49	5.86	5.27	4.78	6.39	4.06	4.74	4.27	3.63	3.36
Volatility (%)	5.41	6.43	6.98	7.37	6.10	7.13	6.84	6.49	6.79	7.15	5.14	6.54	6.27	4.93	4.97	4.63	4.81	4.29	4.40	4.82	4.08	4.63
Sharpe ratio	0.61	0.38	0.39	0.29	0.59	0.49	0.49	0.59	0.56	0.43	_0.64_	0.46	_0.68_	_0.95***_	_0.82_	_0.77_	_1.07***_	_0.66_	_0.80***_	_0.63_	0.59	0.46
Port. turnover (%)	1.60	1.97	2.17	2.31	1.93	2.17	2.09	1.98	2.08	2.22	1.77	2.06	1.94	1.74	1.96	1.75	1.77	1.67	1.49	1.79	1.58	2.14
Avg. com. weight (%)	-	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00
Panel D: Risk-parity																						
Return (%)	3.85	3.84	3.86	3.80	4.02	4.01	4.03	4.20	4.08	3.90	3.72	3.94	4.23	4.30	4.03	3.91	4.59	3.63	3.85	3.67	2.95	3.56
Volatility (%)	2.86	2.92	2.89	2.88	2.90	2.90	2.92	2.96	2.92	2.89	2.89	2.90	2.93	2.80	2.75	2.69	2.77	2.63	2.19	2.79	2.29	2.75
Sharpe ratio	0.92	0.90	0.91	0.89	_0.97_	_0.96_	_0.97_	_1.01*_	_0.98_	_0.93_	0.87	_0.94_	_1.03**_	_1.11***_	_1.02**_	_1.01_	_1.21***_	_0.92_	_1.25*_	0.88	0.76	0.85
Port. turnover (%)	3.06	3.20	3.15	3.14	3.21	3.18	3.18	3.21	3.17	3.14	3.60	3.18	3.26	3.29	3.22	3.40	3.35	3.39	4.09	3.51	4.00	3.21
Avg. com. weight (%)	-	3.19	2.02	1.58	3.67	2.62	2.64	3.68	2.67	1.85	8.10	2.76	4.09	7.30	3.81	7.77	7.26	11.52	49.84	7.24	20.59	2.92
Panel E: Reward-to-risk timing																						
Return (%)	3.90	4.03	4.11	3.94	4.62	4.32	4.68	4.93	4.80	4.25	4.12	4.20	5.23	5.19	4.65	4.77	6.00	3.79	4.43	3.83	3.35	3.90
Volatility (%)	3.37	3.51	3.56	3.45	3.63	3.62	3.75	3.86	3.78	3.57	3.43	3.56	3.96	3.72	3.52	3.53	3.60	3.28	2.47	3.30	2.84	3.25
Sharpe ratio	0.79	_0.80_	_0.81_	0.79	_0.94_	_0.86_	_0.93_	_0.97*_	_0.95_	_0.85_	_0.85_	_0.84_	_1.02*_	_1.08**_	_0.98_	_1.01_	_1.33***_	0.78	_1.35***_	0.79	0.75	_0.82_
Port. turnover (%)	7.74	7.97	7.98	8.00	8.15	7.98	7.94	7.93	7.92	7.95	8.24	8.05	8.02	8.66	8.16	8.09	8.16	8.65	8.69	8.53	9.59	8.06
Avg. com. weight (%)	-	3.90	3.48	2.01	6.03	5.56	5.48	7.21	5.69	3.49	6.96	4.76	8.61	14.24	7.49	11.15	18.71	7.61	43.26	5.79	5.93	2.04

Notes: This table reports the results of several out-sample asset allocation strategies (Panel A to Panel D) for a benchmark (stock–bond) portfolio and the benchmark portfolio complemented with one out of 21 commodity indices at a time. The results are derived from a rolling sample approach with a 36-month estimation window and cover the full sample from February 2000 to April 2017. Return and volatility denote the annualized time series mean and volatility of monthly net returns, respectively. The Sharpe ratio is calculated as annualized time series mean of net excess returns divided by the annualized time series volatility of net excess returns. 'Port. turnover' denotes the portfolio turnover, calculated as the average change in weights over all rebalancing points. 'Avg. com. weight' denotes the average weight in the commodity index over the full sample. Italic font indicates an absolute increase in the Sharpe ratio of the augmented portfolio relative to the benchmark case. *, ** and *** indicate a significant increase in the Sharpe ratio of the augmented portfolio relative to the benchmark case at the 10%, 5% and 1% level, respectively.

Additional Tables

Table A.1: Description of Commodity Indices

Index	Methodology	Constituents	Base Date	Launch Date
First Generation Commodity Indices				
Bloomberg Commodity Index (BCOM)	The index aims at representing the broad market. The relative commodity weights are determined by a 5-year average measure of liquidity and production data. Single commodity weights are restricted between 2% and 15% and related commodity groups are capped at 33%. The index rolls liquid nearby contracts.	Long-only investment in 20 commodities from all sectors.	31.12.1990	14.07.1998
Deutsche Bank Liquid Commodity Index (DBLCI)	The index aims at representing the biggest commodity sectors. The relative weights are fixed and determined by commodity world production and inventories. Energy contracts are rolled monthly, remaining contracts annually.	Long-only investment in six commodities from all sectors excluding livestock.	01.12.1988	28.02.2003
S&P Goldman Sachs Commodity Index (S&P-GSCI)	The index aims at representing the broad market. The relative commodity weights are mainly determined by world production over the last five years. Also, liquidity thresholds for the total dollar value traded are established. The index rolls liquid nearby contracts.	Long-only investment in 24 commodities from all sectors.	31.12.1969	08.01.1991
Second Generation Commodity Indices				
CYD Long Only Commodity Index (CYDLO)	The index is equally weighted invested in all eligible commodity contracts that show backwardation in the actively traded part of the futures term structure. All contract maturities, that fulfil the liquidity requirements, are considered and those with the highest annualized roll returns are selected. Single commodity weights are capped based on liquidity categories at either 5%, 10% or 15%. Commodity sector weights are capped at 50% with the exception of exotics.	The number of long-only invested commodities depends on and varies with the futures term structure of 26 eligible commodities from all sectors.	31.12.1979	October 2006
Deutsche Bank Liquid Commodity Index Mean Reversion (DBLCI-MR)	The index rebalances single commodity weights from predefined base weights depending on the relative richness or cheapness of each commodity. The development of 1- and 5-year moving averages of commodity prices is used to put higher (lower) weights on relative cheap (expensive) commodities.	Long-only investment in six commodities from all sectors excluding livestock.	01.12.1988	14.02.2003
Deutsche Bank Liquid Commodity Index Optimum Yield (DBLCI-OY)	The index weights are almost equal to the DBLCI, representing world production and inventories of the biggest commodity sectors. Different from the DBLCI, this index selects the most backwardated futures contracts. Furthermore, the futures rolling dates are not predefined but rule-based, relying on the implied roll return of eligible contracts that can have maturities up to 13 months.	Long-only investment in six commodities from all sectors excluding livestock.	02.12.1988	31.05.2006
Deutsche Bank Liquid Commodity Index Optimum Yield Balanced (DBLCI-OYBA)	The index follows the methodology of the DBLCI-OY, as outlined above. It differs from the DBLCI-OY by having more index constituents. It differs from the DBLCI-OYBR by having a lower exposure to the energy sector.	Long-only investment in 14 commodities from all sectors excluding livestock.	03.09.1997	11.01.2007
Deutsche Bank Liquid Commodity Index Optimum Yield Broad (DBLCI-OYBR)	The index follows the methodology of the DBLCI-OY, as outlined above. It differs from the DBLCI-OY by having more index constituents. It differs from the DBLCI-OYBA by having a higher exposure to the energy sector.	Long-only investment in 14 commodities from all sectors excluding livestock.	03.09.1997	11.01.2007
Merrill Lynch Commodity Index eXtra (MLCX)	The index aims at representing the broad market. The relative commodity weights are determined by global production and the futures liquidity. Each sector has to be represented by at least two and at most four (five for the energy sector) commodities. Sector weights are capped at 60% and have a lower bound of 3%. The index deviates from the first generation by rolling second- into third-month contracts over an extended roll period of 15 days.	Long-only investment in 20 commodities from all sectors.	29.06.1990	June 2006

Table A.1 Description of Commodity Indices *(continued)*

Index	Methodology	Constituents	Base Date	Launch Date
Second Generation Commodity Indices				
Morningstar Long/Flat Commodity Index (MSLF)	The index is derived from the MSLS and follows the same methodology, as outlined below. It takes the same long positions but differs by replacing all short positions with cash investments.	Investment in up to 20 commodities from all sectors. Long commodity and cash allocations depend on a momentum signal.	21.12.1979	01.08.2007
Morningstar Long-Only Commodity Index (MSLO)	The index is long in all eligible commodity futures contracts that are relevant for the MSLS, as outlined below. It does not take any short or cash positions.	Long-only investment in 20 commodities from all sectors.	21.12.1979	01.08.2007
SummerHaven Dynamic Commodity Index (SDCI)	The index is equally weighted and combines multiple strategies in its monthly rebalancing based on 27 eligible commodities. First, the seven most backwardated commodities are included. Next, from the remaining 20 eligible commodities, those seven with the greatest 12-month price momentum are included. Finally, individual contracts with the highest backwardation are selected while the maximum contract expiration varies for each sector from 5 to 12 months. The index requires each sector to be represented with at least one contract.	Long-only investment in 14 commodities from all sectors.	02.01.1991	December 2009
Third Generation Commodity Indices				
Barclays Backwardation Long Short Index (BBLS)	The index is long in the three-month maturity contracts of the six commodities with the greatest backwardation out of 23 eligible commodities that are represented by Barclays sub-indices. At the same time, the index is short in the nearby or front month contracts of the six commodities with the lowest backwardation.	Long and short investment each in six commodities potentially from all sectors.	08.01.1999	26.11.2010
Credit Suisse Momentum and Volatility Enhanced Return Strategy (CSMOVERS)	The index combines return and volatility signals to equally invest in ten out of 24 eligible commodities that are represented by S&P-GSCI sub-indices. The relation of the short-term commodity futures volatility to its long-term average determines the maturity of selected contracts. If the volatility is relatively low (moderate) [high], contracts with six (three) [one] month[s] maturity are selected. The price momentum determines if long (positive momentum) or short (negative momentum) positions are held. The index has different caps for each commodity sector.	Investment in ten commodities potentially from all sectors. Long or short allocation depends on the commodity futures price momentum.	02.01.1998	15.04.2009
Credit Suisse Momentum and Volatility Enhanced Return Strategy Market Neutral (CSMOVERSMN)	The index follows the methodology of the CSMOVERS, as outlined above, and only differs with regard to the index constituents. The index is equally weighted invested in twelve commodities and is market neutral, i.e. it holds an equal number and same weights of long and short contracts.	Long and short investment each in six commodities potentially from all sectors.	02.01.1998	15.04.2009
CYD Diversified Long/Short Commodity Index (CYDDLS)	The index is equally weighted long invested in ten out of 29 eligible commodities with the greatest backwardation. A diversification rule requires an allocation of at least 10% to every sector and also sets different limits for sectors and individual commodities. The short allocation consists of an equally weighted investment in all 19 remaining commodities. Futures contracts can have maturities up to twelve months.	Long investment in ten commodities and short investment in 19 commodities from all sectors, respectively.	05.01.2000	Spring 2013
CYD Long Short Commodity Index (CYDLS)	The index is equally weighted long invested in all backwardated commodities out of 26 eligible constituents. It is also equally weighted short invested in all contangoed eligible commodities. The futures term structure is only classified as contango if every actively traded maturity is in contango and consequently, the index has a long-bias. All contracts fulfilling the eligibility criteria (e.g. with respect to liquidity) are considered, i.e. there is no strict limit to contract maturity.	Investment in 26 commodities from all sectors. Long or short allocation depends on the commodity futures term structure.	31.12.1979	October 2006

Table A.1 Description of Commodity Indices *(continued)*

Index	Methodology	Constituents	Base Date	Launch Date
Third Generation Commodity Indices				
CYD Market Neutral Plus Commodity Index (CYDMNP)	The index holds a long position in the nearby contract of all eligible commodities and short positions in the 2nd- and 3rd-nearby contracts of the same commodity, given sufficient futures contract liquidity. The index has a leverage factor of two, i.e. for every dollar invested, the index engages in long and short positions of the same commodity each with a notional of one dollar. The relative commodity weights follow predefined target weights ranging from 3% to 9% for each of 15 commodities.	Long and short investment each in 15 commodities from all sectors.	31.12.1979	October 2006
Morningstar Long/Short Commodity Index (MSLS)	The index follows a momentum rule to determine investment positions in commodity futures. Therefore, a linked price series for each commodity is computed and related to the commodities' 12-month moving average. If the linked price exceeds (falls below) the 12-month moving average, a long (short) position is taken. An exception is made for the energy sector, where short positions are replaced by a cash investment (flat position). The magnitude of this momentum signal determines the relative weight of each commodity. Individual contracts are capped at 10%. The index considers all eligible commodity futures contracts that rank in the top 95% of the total dollar value of open interest. Eligibility requires trading on U.S. exchanges and price denominated in U.S. dollars.	Investment in 20 commodities from all sectors. Long or short allocation depends on a momentum signal.	21.12.1979	01.08.2007
Morningstar Short/Flat Commodity Index (MSSF)	The index is derived from the MSLS and follows the same methodology, as outlined above. It takes the same short positions but differs by replacing all long positions with cash investments.	Investment in up to 20 commodities from all sectors. Short commodity and cash allocations depends on a momentum signal.	21.12.1979	01.08.2007
Morningstar Short-Only Commodity Index (MSSO)	The index is short in all eligible commodity futures contracts that are relevant for the MSLS, as outlined above. It does not take any long or cash positions.	Short-only investment in 20 commodities from all sectors.	21.12.1979	01.08.2007

Notes: This table describes the set of commodity indices used in this study. Column two, 'Methodology', refers to the rules and guidelines that determine the index constituents, weights and whether positions are entered long or short. Column three, 'Constituents', summarizes in how many commodity futures an index is invested, which sectors are covered and whether long and/or short positions are taken. Column four, 'Base Date', names the date to which the index is calculated back to, potentially using backfilled data. Column five, 'Launch Date', shows the date when the index was first publicly available and calculated with live market data. The information presented in this table is taken from the index providers' websites.

Table A.2: Results of Sub-Sample Spanning Tests for Commodity Indices (December 2010 – April 2017)

Commodity	Wald	LR	LM	GMM-Wald	F_1	F_2
First Generation Commodity Indices						
BCOM	11.066***	10.340***	9.676***	9.629***	7.526***	2.860*
	(0.004)	(0.006)	(0.008)	(0.008)	(0.008)	(0.095)
DBLCI	10.199***	9.578***	9.006**	9.576***	5.544**	4.014**
	(0.006)	(0.008)	(0.011)	(0.008)	(0.021)	(0.049)
S&P-GSCI	13.104***	12.101***	11.198***	13.196***	4.792**	7.426***
	(0.001)	(0.002)	(0.004)	(0.001)	(0.032)	(0.008)
Second Generation Commodity Indices						
CYDLO	17.502***	15.771***	14.261***	14.074***	4.042**	12.281***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.048)	(0.001)
DBLCI-MR	8.095**	7.697**	7.325**	8.799**	4.113**	3.520*
	(0.017)	(0.021)	(0.026)	(0.012)	(0.046)	(0.065)
DBLCI-OY	11.835***	11.009***	10.258***	11.157***	6.402**	4.637**
	(0.003)	(0.004)	(0.006)	(0.004)	(0.014)	(0.035)
DBLCI-OYBA	8.242**	7.830**	7.445**	6.823**	5.665**	2.125
	(0.016)	(0.020)	(0.024)	(0.033)	(0.020)	(0.149)
DBLCI-OYBR	9.857***	9.275***	8.739**	8.648**	4.973**	4.273**
	(0.007)	(0.010)	(0.013)	(0.013)	(0.029)	(0.042)
MLCX	10.912***	10.205***	9.558***	11.011***	4.442**	5.780**
	(0.004)	(0.006)	(0.008)	(0.004)	(0.038)	(0.019)
MSLF	12.227***	11.348***	10.551***	11.852***	4.225**	7.215***
	(0.002)	(0.003)	(0.005)	(0.003)	(0.043)	(0.009)
MSLO	8.848**	8.375**	7.936**	7.777**	4.383**	3.942*
	(0.012)	(0.015)	(0.019)	(0.020)	(0.040)	(0.051)
SDCI	10.207***	9.585***	9.013**	9.182**	5.029**	4.537**
	(0.006)	(0.008)	(0.011)	(0.010)	(0.028)	(0.036)
Third Generation Commodity Indices (CYDDLS removed)						
BBLS	16.272***	14.762***	13.433***	21.219***	0.597	15.122***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.442)	(0.000)
CSMOVERS	7.565**	7.216**	6.888**	7.704**	0.017	7.349***
	(0.023)	(0.027)	(0.032)	(0.021)	(0.896)	(0.008)
CSMOVERSMN	17.798***	16.012***	14.457***	17.343***	1.264	15.785***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.265)	(0.000)
CYDLS	45.647***	35.844***	28.658***	50.789***	0.030	44.414***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.864)	(0.000)
CYDMNP	118.951***	71.923***	46.742***	133.993***	0.204	115.338***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.653)	(0.000)
MSLS	16.054***	14.582***	13.285***	16.670***	1.418	13.934***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.238)	(0.000)
MSSF	43.759***	34.649***	27.902***	31.557***	1.877	39.713***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.175)	(0.000)
MSSO	5.281*	5.108*	4.942*	6.023**	3.169*	1.853
	(0.071)	(0.078)	(0.084)	(0.049)	(0.079)	(0.178)

Notes: This table reports test statistics and respective p-values (in parentheses) for several tests of the null hypotheses that stocks and bonds span commodities. Columns two to four present the Wald, Likelihood ratio (LR) and Lagrange multiplier (LM) tests. The fifth column reports the results of the Wald test based on the GMM estimation (GMM-Wald). Columns six and seven report the results of the step-down procedure: F_1 tests for a return enhancement and F_2 for a risk reduction. Stocks are represented by the S&P 500 Index, bonds by the Bloomberg Barclays US Aggregate Bond Index and commodities by the respective index presented in each row. The analysis is based on monthly excess returns over the 1-month U.S. Treasury Bill and covers the period from December 2010 to April 2017. *, ** and *** indicate significant entries at the 10%, 5% and 1% significance level, respectively.

Table A.3: Spanning Tests of Equally Sized Sub-Samples

	BCOM	DBLCI	SPGSCI	CYDLO	DBLCIMR	DBLCI-OY	DBLCI-OYBA	DBLCI-OYBR	MLCX	MSLF	MSLO	SDCI	BBLS	CSMOVERS	CSMOVERSMN	CYDDLS	CYDLS	CYDMNP	MSLS	MSSF	MSSO
Panel A: Period February 2000–April 2004																					
Wald	2.02	1.67	1.11	4.06	2.77	4.60	4.69*	5.39*	3.98	3.48	2.66	7.06**	7.09**	3.24	9.23***	20.30***	2.95	36.78***	4.09	61.09***	14.72***
p-value	0.36	0.43	0.58	0.13	0.25	0.10	0.10	0.07	0.14	0.18	0.27	0.03	0.03	0.20	0.01	0.00	0.23	0.00	0.13	0.00	0.00
Panel B: Period May 2004 - August 2008																					
Wald	5.72*	7.87**	4.88*	10.70***	9.36***	10.74***	11.86***	11.16***	6.79**	6.62**	6.90**	15.95***	21.44***	9.17**	15.65***	17.06***	7.82**	69.93***	5.60*	22.26***	2.56
p-value	0.06	0.02	0.09	0.00	0.01	0.00	0.00	0.00	0.03	0.04	0.03	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.06	0.00	0.28
Panel C: Period September 2008 - December 2012																					
Wald	3.45	3.19	4.79*	2.19	3.58	2.10	1.16	1.33	2.42	9.21**	1.08	1.44	0.95	10.60***	11.14***	7.37**	14.29***	101.82***	30.76***	90.71***	15.65***
p-value	0.18	0.20	0.09	0.33	0.17	0.35	0.56	0.51	0.30	0.01	0.58	0.49	0.62	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Panel D: Period January 2013 - April 2017																					
Wald	1.91	0.93	0.84	1.51	0.41	0.69	1.70	0.85	0.68	6.13**	1.49	1.98	7.61**	5.81*	29.96***	9.56***	7.78**	69.59***	4.71*	4.54	0.62
p-value	0.38	0.63	0.66	0.47	0.82	0.71	0.43	0.65	0.71	0.05	0.47	0.37	0.02	0.05	0.00	0.01	0.02	0.00	0.09	0.10	0.73
Panel E: Number of rejections (at the 10% level)																					
	1	1	2	1	1	1	2	2	1	3	1	2	3	3	4	4	3	4	3	3	2

Notes: This table reports test statistics and respective p-values of a Wald test of the null hypotheses that stocks and bonds span commodities. Panel A to D present the results of four equally sized sub-samples that partition the February 2000 to April 2017 period. Panel E lists how many times the null of spanning is rejected for each commodity index at the 10% significance level. Stocks are represented by the S&P 500 Index, bonds by the Bloomberg Barclays US Aggregate Bond Index and commodities by the respective index presented in each column. The analysis is based on monthly excess returns over the 1-month U.S. Treasury Bill.

Table A.4: Results of Spanning Tests for Commodity Indices with Extended Benchmark Assets (February 2000 – April 2017)

Commodity	Wald	LR	LM	GMM-Wald	F_1	F_2
First Generation Commodity Indices						
BCOM	0.711	0.710	0.709	0.723	0.626	0.068
	(0.701)	(0.701)	(0.702)	(0.697)	(0.430)	(0.794)
DBLCI	0.099	0.099	0.099	0.087	0.025	0.072
	(0.952)	(0.952)	(0.952)	(0.957)	(0.875)	(0.789)
S&P-GSCI	0.840	0.839	0.837	0.805	0.293	0.529
	(0.657)	(0.658)	(0.658)	(0.669)	(0.589)	(0.468)
Second Generation Commodity Indices						
CYDLO	1.565	1.560	1.554	1.001	1.068	0.460
	(0.457)	(0.458)	(0.460)	(0.606)	(0.303)	(0.499)
DBLCI-MR	0.276	0.276	0.276	0.192	0.080	0.191
	(0.871)	(0.871)	(0.871)	(0.908)	(0.778)	(0.663)
DBLCI-OY	0.176	0.176	0.176	0.117	0.129	0.043
	(0.916)	(0.916)	(0.916)	(0.943)	(0.720)	(0.836)
DBLCI-OYBA	0.385	0.385	0.384	0.394	0.356	0.020
	(0.825)	(0.825)	(0.825)	(0.821)	(0.552)	(0.887)
DBLCI-OYBR	0.676	0.675	0.674	0.524	0.633	0.026
	(0.713)	(0.714)	(0.714)	(0.769)	(0.427)	(0.871)
MLCX	0.218	0.218	0.217	0.141	0.146	0.066
	(0.897)	(0.897)	(0.897)	(0.932)	(0.702)	(0.797)
MSLF	8.597**	8.423**	8.254**	12.909***	1.781	6.583**
	(0.014)	(0.015)	(0.016)	(0.002)	(0.184)	(0.011)
MSLO	0.135	0.135	0.135	0.103	0.008	0.125
	(0.935)	(0.935)	(0.935)	(0.950)	(0.931)	(0.724)
SDCI	2.661	2.644	2.627	2.111	2.558	0.038
	(0.264)	(0.267)	(0.269)	(0.348)	(0.111)	(0.845)
Third Generation Commodity Indices						
BBLS	16.987***	16.326***	15.699***	15.734***	12.863***	3.509*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.062)
CSMOVERS	19.943***	19.040***	18.190***	9.163**	7.341***	11.753***
	(0.000)	(0.000)	(0.000)	(0.010)	(0.007)	(0.001)
CSMOVERSMN	30.638***	28.572***	26.688***	25.956***	7.247***	21.975***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.008)	(0.000)
CYDDLS	33.471***	31.025***	28.812***	25.485***	26.165***	5.781**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.017)
CYDLS	29.951***	27.973***	26.165***	16.940***	2.024	27.067***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.156)	(0.000)
CYDMNP	153.065***	114.588***	87.996***	188.026***	25.468***	110.572***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MSLS	19.373***	18.519***	17.715***	22.001***	3.080*	15.664***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.081)	(0.000)
MSSF	140.648***	107.323***	83.746***	103.602***	1.013	136.228***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.315)	(0.000)
MSSO	21.329***	20.301***	19.337***	18.706***	0.003	20.914***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.954)	(0.000)

Notes: This table reports test statistics and respective p-values (in parentheses) for several tests of the null hypotheses that stocks, bonds, value and size equity portfolios span commodities. Columns two to four present the Wald, Likelihood ratio (LR) and Lagrange multiplier (LM) tests. The fifth column reports the results of the Wald test based on the GMM estimation (GMM-Wald). Columns six and seven report the results of the step-down procedure: F_1 tests for a return enhancement and F_2 for a risk reduction. Stocks are represented by the S&P 500 Index, bonds by the Bloomberg Barclays US Aggregate Bond Index, value and size portfolio returns by the HML and SMB data from Kenneth French's website and commodities by the respective index presented in each row. The analysis is based on monthly excess returns over the 1-month U.S. Treasury Bill and covers the period from February 2000 to April 2017. *, ** and *** indicate significant entries at the 10%, 5% and 1% significance level, respectively.

Table A.5: Results of Sub-Sample Spanning Tests for Commodity Indices with Extended Benchmark Assets (December 2010 – April 2017)

Commodity	Wald	LR	LM	GMM-Wald	F_1	F_2
First Generation Commodity Indices						
BCOM	9.440***	8.905**	8.409**	8.067**	8.795***	0.029
	(0.009)	(0.012)	(0.015)	(0.018)	(0.004)	(0.865)
DBLCI	7.466**	7.126**	6.806**	6.590**	6.969**	0.011
	(0.024)	(0.028)	(0.033)	(0.037)	(0.010)	(0.915)
S&P-GSCI	7.496**	7.154**	6.831**	6.822**	6.128**	0.824
	(0.024)	(0.028)	(0.033)	(0.033)	(0.016)	(0.367)
Second Generation Commodity Indices						
CYDLO	9.468***	8.930**	8.432**	9.284***	4.601**	4.053**
	(0.009)	(0.012)	(0.015)	(0.010)	(0.035)	(0.048)
DBLCI-MR	6.600**	6.332**	6.079**	5.843*	5.803**	0.346
	(0.037)	(0.042)	(0.048)	(0.054)	(0.019)	(0.558)
DBLCI-OY	8.727**	8.267**	7.838**	7.416**	8.133***	0.025
	(0.013)	(0.016)	(0.020)	(0.025)	(0.006)	(0.875)
DBLCI-OYBA	7.486**	7.144**	6.823**	6.231**	6.937**	0.059
	(0.024)	(0.028)	(0.033)	(0.044)	(0.010)	(0.809)
DBLCI-OYBR	6.967**	6.670**	6.389**	6.354**	6.494**	0.019
	(0.031)	(0.036)	(0.041)	(0.042)	(0.013)	(0.891)
MLCX	6.510**	6.249**	6.002**	6.045**	5.799**	0.270
	(0.039)	(0.044)	(0.050)	(0.049)	(0.019)	(0.605)
MSLF	7.322**	6.994**	6.686**	8.190**	3.993**	2.740
	(0.026)	(0.030)	(0.035)	(0.017)	(0.049)	(0.102)
MSLO	6.090**	5.861*	5.644*	5.499*	5.408**	0.270
	(0.048)	(0.053)	(0.059)	(0.064)	(0.023)	(0.605)
SDCI	7.884**	7.506**	7.152**	7.639**	5.534**	1.731
	(0.019)	(0.023)	(0.028)	(0.022)	(0.021)	(0.192)
Third Generation Commodity Indices						
BBLS	17.946***	16.132***	14.554***	14.123***	0.873	15.936***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.353)	(0.000)
CSMOVERS	5.171*	5.005*	4.845*	5.037*	0.000	4.902**
	(0.075)	(0.082)	(0.089)	(0.081)	(0.988)	(0.030)
CSMOVERSMN	9.972***	9.377***	8.829**	10.724***	1.052	8.267***
	(0.007)	(0.009)	(0.012)	(0.005)	(0.309)	(0.005)
CYDLS	42.932***	34.120***	27.564***	44.976***	0.188	40.406***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.666)	(0.000)
CYDMNP	68.041***	48.757***	36.122***	83.055***	0.308	63.921***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.581)	(0.000)
MSLS	10.503***	9.846***	9.243***	15.326***	1.104	8.704***
	(0.005)	(0.007)	(0.010)	(0.000)	(0.297)	(0.004)
MSSF	32.872***	27.375***	23.037***	26.260***	2.231	28.034***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.140)	(0.000)
MSSO	10.148***	9.533***	8.966**	8.350**	4.303**	4.961**
	(0.006)	(0.009)	(0.011)	(0.015)	(0.042)	(0.029)

Notes: This table reports test statistics and respective p-values (in parentheses) for several tests of the null hypotheses that stocks, bonds, value and size equity portfolios span commodities. Columns two to four present the Wald, Likelihood ratio (LR) and Lagrange multiplier (LM) tests. The fifth column reports the results of the Wald test based on the GMM estimation (GMM-Wald). Columns six and seven report the results of the step-down procedure: F_1 tests for a return enhancement and F_2 for a risk reduction. Stocks are represented by the S&P 500 Index, bonds by the Bloomberg Barclays US Aggregate Bond Index, value and size portfolio returns by the HML and SMB data from Kenneth French's website and commodities by the respective index presented in each row. The analysis is based on monthly excess returns over the 1-month U.S. Treasury Bill and covers the period from December 2010 to April 2017. *, ** and *** indicate significant entries at the 10%, 5% and 1% significance level, respectively.

Table A.6: Portfolio Optimization with Increased Transaction Costs

	Benchmark	BCOM	DBLCI	S&P-GSCI	CYDLO	DBLCI-MR	DBLCI-OY	DBLCI-OYBA	DBLCI-OYBR	MLCX	MSLF	MSLO	SDCI	BBLS	CSMOVERS	CSMOVERSMN	CYDDLS	CYDLS	CYDMNP	MSLS	MSSF	MSSO
		Benchmark portfolio complemented with commodity index																				
Panel A: mv (aggressive)																						
Return (%)	6.67	8.22	9.38	7.94	11.96	11.92	12.91	12.00	12.88	10.66	7.13	9.06	14.27	12.93	12.06	10.31	16.31	6.06	8.34	5.17	5.74	4.18
Volatility (%)	8.51	11.19	12.75	10.73	12.37	13.20	13.42	12.74	13.33	12.23	10.71	12.25	13.00	9.90	12.44	9.52	9.98	8.48	7.07	9.94	8.49	8.76
Sharpe ratio	0.64	0.62	_0.64_	0.62	_0.86_	_0.81_	_0.87_	_0.84_	_0.87_	_0.77_	0.55	_0.64_	_1.00_	_1.18**_	_0.87_	_0.95_	_1.50***_	0.57	_1.00**_	0.40	0.53	0.34
Port. turnover (%)	14.12	14.79	12.32	12.62	8.47	12.57	8.67	9.77	9.47	11.15	20.46	13.32	11.42	13.80	13.83	13.26	13.47	20.71	16.43	22.78	17.69	20.69
Avg. com. weight (%)	-	22.12	26.53	14.38	32.64	34.71	32.97	33.49	31.84	25.69	28.76	28.13	41.03	42.28	40.49	36.20	68.59	16.02	29.83	22.35	2.03	9.64
Panel B: mv (conservative)																						
Return (%)	4.59	5.30	5.61	5.16	6.45	6.03	6.49	6.66	6.63	6.02	5.29	5.75	7.36	7.03	6.61	6.70	9.44	4.30	6.76	4.25	3.88	4.64
Volatility (%)	4.60	4.85	5.34	4.79	5.25	5.08	5.39	5.39	5.41	5.07	4.97	5.13	5.51	4.88	5.18	4.93	5.34	4.73	3.60	4.76	4.46	4.72
Sharpe ratio	0.72	_0.83_	_0.82_	_0.82_	_0.99*_	_0.95_	_0.98_	_1.01*_	_1.00*_	_0.94_	_0.81_	_0.88_	_1.12*_	_1.20**_	_1.04_	_1.12**_	_1.54***_	0.65	_1.56***_	0.63	0.60	0.72
Port. turnover (%)	12.36	11.15	10.80	11.22	10.18	11.57	10.14	10.33	10.16	9.99	13.25	10.87	11.03	14.74	12.36	12.64	8.96	16.16	10.52	14.60	15.41	14.52
Avg. com. weight (%)	-	6.81	7.03	4.04	9.96	9.80	9.40	10.99	9.22	6.81	12.29	8.37	14.21	20.39	14.81	20.66	36.38	11.14	35.56	10.54	4.80	5.57
Panel C: 1/N with rebalancing																						
Return (%)	4.50	3.60	3.89	3.24	4.77	4.60	4.51	4.97	4.92	4.23	4.45	4.19	5.44	5.82	5.22	4.73	6.34	4.02	4.71	4.22	3.59	3.31
Volatility (%)	5.41	6.43	6.98	7.37	6.10	7.13	6.85	6.49	6.79	7.15	5.14	6.54	6.27	4.93	4.97	4.63	4.81	4.29	4.40	4.82	4.08	4.63
Sharpe ratio	0.61	0.37	0.39	0.28	0.58	0.48	0.48	0.58	0.55	0.42	_0.63_	0.46	_0.68_	_0.94***_	_0.81_	_0.77_	_1.06***_	_0.65_	_0.79***_	_0.62_	0.58	0.45
Port. turnover (%)	1.60	1.97	2.17	2.31	1.93	2.17	2.09	1.98	2.08	2.22	1.77	2.06	1.94	1.74	1.96	1.75	1.77	1.67	1.49	1.79	1.58	2.14
Avg. com. weight (%)	-	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00
Panel D: Risk-parity																						
Return (%)	3.78	3.76	3.78	3.72	3.94	3.93	3.95	4.12	4.00	3.82	3.63	3.86	4.15	4.22	3.95	3.82	4.51	3.55	3.75	3.58	2.85	3.48
Volatility (%)	2.86	2.93	2.90	2.88	2.90	2.90	2.92	2.96	2.92	2.89	2.89	2.90	2.93	2.80	2.75	2.69	2.77	2.63	2.19	2.79	2.29	2.75
Sharpe ratio	0.89	0.87	0.89	0.87	_0.94_	_0.94_	_0.94_	_0.98*_	_0.95_	_0.90_	0.84	_0.91_	_1.00**_	_1.08***_	_0.99*_	_0.98_	_1.18***_	0.89	_1.21*_	0.85	0.71	0.82
Port. turnover (%)	3.06	3.20	3.15	3.14	3.21	3.18	3.18	3.21	3.17	3.14	3.60	3.18	3.26	3.29	3.22	3.40	3.35	3.39	4.09	3.51	4.00	3.21
Avg. com. weight (%)	-	3.19	2.02	1.58	3.67	2.62	2.64	3.68	2.67	1.85	8.10	2.76	4.09	7.30	3.81	7.77	7.26	11.52	49.84	7.24	20.59	2.92
Panel E: Reward-to-risk timing																						
Return (%)	3.71	3.84	3.91	3.75	4.41	4.12	4.49	4.73	4.60	4.05	3.91	4.00	5.03	4.98	4.44	4.57	5.80	3.57	4.21	3.61	3.11	3.70
Volatility (%)	3.37	3.51	3.56	3.45	3.62	3.63	3.75	3.86	3.78	3.57	3.43	3.56	3.96	3.72	3.52	3.53	3.60	3.27	2.47	3.30	2.85	3.25
Sharpe ratio	0.74	_0.75_	_0.76_	0.73	_0.88_	_0.81_	_0.88_	_0.92*_	_0.90_	_0.79_	_0.79_	_0.78_	_0.97*_	_1.02**_	_0.92_	_0.96_	_1.27***_	0.72	_1.26***_	0.73	0.67	_0.76_
Port. turnover (%)	7.74	7.97	7.98	8.00	8.15	7.98	7.94	7.93	7.92	7.95	8.24	8.05	8.02	8.66	8.16	8.09	8.16	8.65	8.69	8.53	9.59	8.06
Avg. com. weight (%)	-	3.90	3.48	2.01	6.03	5.56	5.48	7.21	5.69	3.49	6.96	4.76	8.61	14.24	7.49	11.15	18.71	7.61	43.26	5.79	5.93	2.04

Notes: This table reports the results of several out-of-sample asset allocation strategies (Panel A to Panel D) for a benchmark (stock–bond) portfolio and the benchmark portfolio complemented with one out of 21 commodity indices at a time. The results are derived from a rolling sample approach with a 36-month estimation window increased transaction costs of 50 basis points and cover the full sample from February 2000 to April 2017. Return and volatility denote the annualized time series mean and volatility of monthly net returns, respectively. The Sharpe ratio is calculated as annualized time series mean of net excess returns divided by the annualized time series volatility of net excess returns. 'Port. turnover' denotes the portfolio turnover, calculated as the average change in weights over all rebalancing points. 'Avg. com. weight' denotes the average weight in the commodity index over the full sample. Italic font indicates an absolute increase in the Sharpe ratio of the augmented portfolio relative to the benchmark case. *, ** and *** indicate a significant increase in the Sharpe ratio of the augmented portfolio relative to the benchmark case at the 10%, 5% and 1% level, respectively.

Table A.7: Portfolio Optimization with 48-Month Estimation Window

	Benchmark	Benchmark portfolio complemented with commodity index																				
	Benchmark	BCOM	DBLCI	S&P-GSCI	CYDLO	DBLCI-MR	DBLCI-OY	DBLCI-OYBA	DBLCI-OYBR	MLCX	MSLF	MSLO	SDCI	BBLS	CSMOVERS	CSMOVERSMN	CYDDLS	CYDLS	CYDMNP	MSLS	MSSF	MSSO
Panel A: mv (aggressive)																						
Return (%)	5.88	6.60	8.74	6.27	11.16	11.64	11.20	11.59	11.95	9.21	8.67	8.04	15.35	15.11	12.85	11.49	18.14	7.59	9.55	7.85	5.76	4.01
Volatility (%)	8.68	11.74	12.70	11.24	12.62	13.82	13.62	13.52	13.74	12.80	10.90	12.92	13.91	10.06	12.89	9.69	9.86	8.12	6.47	10.27	8.66	8.57
Sharpe ratio	0.53	0.46	_0.59_	0.45	_0.78_	_0.75_	_0.73_	_0.76_	_0.78_	_0.62_	_0.68_	0.53	_1.01*_	_1.38***_	_0.90_	_1.06**_	_1.71***_	_0.78_	_1.28***_	_0.64_	0.52	0.33
Port. turnover (%)	16.26	16.39	12.86	14.83	10.54	10.10	10.96	10.58	11.48	9.91	19.62	15.54	7.87	10.04	18.07	16.45	5.84	15.95	15.66	20.04	17.34	20.69
Avg. com. weight (%)	-	21.34	26.34	14.22	38.11	41.20	36.54	40.19	37.93	26.91	36.57	29.96	48.31	54.83	49.88	45.74	70.12	18.08	36.06	30.98	0.37	2.64
Panel B: mv (conservative)																						
Return (%)	3.95	4.03	4.63	4.03	5.62	5.45	5.62	6.24	5.87	4.68	4.84	4.49	7.14	7.30	6.18	6.58	9.61	4.42	6.89	4.33	3.49	3.93
Volatility (%)	4.53	5.15	5.38	5.13	5.33	5.17	5.49	5.58	5.47	5.34	5.03	5.44	5.74	4.74	4.99	4.81	5.17	4.56	3.37	4.68	4.50	4.70
Sharpe ratio	0.59	0.54	_0.63_	0.54	_0.82_	_0.82_	_0.80_	_0.90*_	_0.85_	_0.64_	_0.71_	_0.60_	_1.03**_	_1.30***_	_0.99*_	_1.12**_	_1.62***_	_0.69_	_1.71***_	_0.66_	0.50	0.57
Port. turnover (%)	9.78	10.12	9.34	9.74	8.59	9.20	8.80	8.81	8.62	8.81	11.25	9.54	8.22	12.09	10.66	10.09	7.57	11.15	8.46	11.20	12.79	11.79
Avg. com. weight (%)	-	7.25	7.04	4.39	11.26	11.70	10.29	13.07	10.56	6.84	14.65	8.83	15.34	23.59	16.50	23.11	37.36	13.76	40.52	12.67	2.81	4.58
Panel C: 1/N with rebalancing																						
Return (%)	3.64	2.67	2.95	2.36	3.93	3.70	3.60	4.12	4.05	3.27	3.72	3.30	4.62	5.34	4.60	4.14	5.74	3.31	3.96	3.57	3.00	2.95
Volatility (%)	5.44	6.46	7.00	7.40	6.12	7.19	6.90	6.55	6.86	7.20	5.14	6.58	6.32	4.99	4.96	4.64	4.86	4.26	4.39	4.81	4.11	4.71
Sharpe ratio	0.44	0.23	0.25	0.16	_0.44_	0.35	0.35	_0.44_	0.41	0.29	_0.49_	0.32	_0.54_	_0.83***_	_0.69*_	_0.63*_	_0.93***_	_0.49_	_0.62***_	_0.49_	0.43	0.37
Port. turnover (%)	1.61	2.00	2.18	2.32	1.94	2.19	2.12	2.01	2.12	2.24	1.78	2.09	1.96	1.75	1.96	1.75	1.77	1.67	1.49	1.80	1.59	2.15
Avg. com. weight (%)	-	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00
Panel D: Risk-parity																						
Return (%)	3.27	3.26	3.30	3.23	3.47	3.46	3.48	3.67	3.53	3.32	3.20	3.36	3.70	3.86	3.52	3.46	4.16	2.99	3.52	3.14	2.71	3.09
Volatility (%)	2.77	2.82	2.80	2.78	2.78	2.81	2.83	2.87	2.83	2.80	2.78	2.80	2.83	2.73	2.64	2.58	2.67	2.51	1.99	2.70	2.25	2.69
Sharpe ratio	0.73	0.72	_0.74_	0.72	_0.80_	_0.79_	_0.79_	_0.85**_	_0.81*_	_0.75_	0.71	_0.76_	_0.87**_	_0.96***_	_0.87**_	_0.87*_	_1.09***_	0.70	_1.21**_	0.71	0.65	0.69
Port. turnover (%)	2.56	2.67	2.66	2.64	2.67	2.65	2.67	2.69	2.67	2.64	2.97	2.67	2.71	2.76	2.67	2.79	2.78	2.83	3.18	2.84	2.94	2.69
Avg. com. weight (%)	-	3.16	2.03	1.59	3.64	2.58	2.59	3.55	2.62	1.86	7.17	2.77	3.93	7.39	3.77	7.69	7.38	11.64	50.22	6.67	20.05	2.90
Panel E: Reward-to-risk timing																						
Return (%)	3.51	3.58	3.71	3.50	4.18	4.14	4.30	4.64	4.40	3.75	3.87	3.72	5.14	5.18	4.50	4.60	6.14	3.51	4.29	3.71	3.06	3.38
Volatility (%)	3.06	3.22	3.26	3.16	3.35	3.35	3.49	3.67	3.54	3.27	3.09	3.27	3.77	3.43	3.00	3.04	3.36	2.98	2.25	2.96	2.77	3.02
Sharpe ratio	0.74	0.72	_0.76_	0.71	_0.88_	_0.87_	_0.88_	_0.93_	_0.90_	_0.76_	_0.85_	_0.76_	_1.05**_	_1.16**_	_1.09**_	_1.11**_	_1.45***_	_0.76_	_1.41***_	_0.83_	0.65	0.70
Port. turnover (%)	6.46	6.67	6.65	6.64	6.64	6.54	6.55	6.49	6.53	6.62	6.79	6.70	6.51	6.94	6.47	6.58	6.52	7.17	7.16	6.88	7.84	6.66
Avg. com. weight (%)	-	3.79	3.53	2.02	6.72	5.89	5.85	7.76	6.15	3.59	7.52	5.00	9.27	15.80	7.93	11.92	20.39	7.51	43.97	6.05	5.22	1.69

Notes: This table reports the results of several out-sample asset allocation strategies (Panel A to Panel D) for a benchmark (stock–bond) portfolio and the benchmark portfolio complemented with one out of 21 commodity indices at a time. The results are derived from a rolling sample approach with a 48-month estimation window and cover the full sample from February 2000 to April 2017. Return and volatility denote the annualized time series mean and volatility of monthly net returns, respectively. The Sharpe ratio is calculated as annualized time series mean of net excess returns divided by the annualized time series volatility of net excess returns. 'Port. turnover' denotes the portfolio turnover, calculated as the average change in weights over all rebalancing points. 'Avg. com. weight' denotes the average weight in the commodity index over the full sample. Italic font indicates an absolute increase in the Sharpe ratio of the augmented portfolio relative to the benchmark case. *, ** and *** indicate a significant increase in the Sharpe ratio of the augmented portfolio relative to the benchmark case at the 10%, 5% and 1% level, respectively.

Table A.8: Portfolio Optimization with 60-Month Estimation Window

	Benchmark	BCOM	DBLCI	S&P-GSCI	CYDLO	DBLCI-MR	DBLCI-OY	DBLCI-OYBA	DBLCI-OYBR	MLCX	MSLF	MSLO	SDCI	BBLS	CSMOVERS	CSMOVERSMN	CYDDLS	CYDLS	CYDMNP	MSLS	MSSF	MSSO
		Benchmark portfolio complemented with commodity index																				
Panel A: mv (aggressive)																						
Return (%)	4.17	4.30	6.27	3.63	8.81	11.29	9.33	10.80	10.44	6.00	4.94	5.89	13.47	14.45	12.08	10.67	17.19	4.68	7.51	2.62	4.16	2.92
Volatility (%)	8.50	11.68	12.95	11.64	12.90	14.88	14.49	14.28	14.55	12.98	10.53	13.06	14.67	10.03	13.31	9.87	10.19	7.99	6.21	9.56	8.50	8.34
Sharpe ratio	0.35	0.26	_0.39_	0.21	_0.59_	_0.68_	_0.56_	_0.67_	_0.63_	_0.37_	_0.35_	_0.36_	_0.83*_	_1.32***_	_0.82_	_0.96**_	_1.56***_	_0.43_	_1.01***_	0.15	0.34	0.21
Port. turnover (%)	12.28	11.15	9.96	12.33	8.39	8.18	8.41	8.61	8.46	8.24	17.90	11.42	5.78	5.14	9.16	8.62	3.03	14.21	8.11	23.71	12.34	15.09
Avg. com. weight (%)	-	21.39	25.58	15.31	40.69	41.88	39.92	45.78	41.74	25.51	47.04	31.37	51.23	64.07	53.63	53.51	72.82	18.87	41.96	32.06	0.01	1.96
Panel B: mv (conservative)																						
Return (%)	3.19	3.40	3.84	3.17	4.68	5.26	4.74	5.61	5.08	3.76	3.72	3.74	6.47	6.90	5.54	5.67	8.26	3.66	5.81	3.14	3.02	3.19
Volatility (%)	4.60	5.24	5.33	5.34	5.29	5.43	5.55	5.68	5.55	5.29	4.73	5.35	5.72	4.79	4.82	4.85	5.33	4.45	3.30	4.24	4.57	4.79
Sharpe ratio	0.43	0.42	_0.49_	0.37	_0.66_	_0.75*_	_0.64_	_0.78*_	_0.70_	_0.48_	_0.53_	_0.47_	_0.93**_	_1.20***_	_0.90*_	_0.93**_	_1.32***_	_0.54_	_1.42***_	_0.45_	0.39	0.41
Port. turnover (%)	8.49	8.22	7.82	8.28	7.11	7.15	7.07	6.95	6.92	7.50	9.57	8.32	6.61	8.84	8.55	7.90	6.69	9.20	5.99	10.09	9.49	9.95
Avg. com. weight (%)	-	6.80	6.81	4.06	11.55	12.24	10.78	13.83	11.21	6.67	16.49	8.64	15.50	26.87	17.78	25.04	38.91	14.23	46.27	13.47	0.89	3.71
Panel C: 1/N with rebalancing																						
Return (%)	3.56	2.45	2.46	1.84	3.51	3.35	2.99	3.71	3.46	2.72	3.40	2.97	4.28	5.09	4.32	3.86	5.45	2.95	3.70	3.19	2.86	3.08
Volatility (%)	5.56	6.64	7.18	7.59	6.28	7.42	7.08	6.73	7.04	7.38	5.24	6.77	6.50	5.07	5.00	4.71	5.00	4.34	4.47	4.86	4.17	4.75
Sharpe ratio	0.42	0.19	0.18	0.09	0.37	0.29	0.25	0.37	0.32	0.21	0.42	0.26	_0.47_	_0.77***_	_0.63_	_0.57_	_0.84***_	0.40	_0.56***_	0.41	0.40	0.39
Port. turnover (%)	1.65	2.03	2.20	2.32	1.96	2.24	2.14	2.05	2.13	2.24	1.79	2.12	1.99	1.77	1.99	1.77	1.79	1.69	1.52	1.80	1.62	2.21
Avg. com. weight (%)	-	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00
Panel D: Risk-parity																						
Return (%)	3.19	3.16	3.14	3.08	3.31	3.35	3.28	3.48	3.33	3.16	3.09	3.23	3.58	3.76	3.41	3.34	4.01	2.84	3.10	2.98	2.61	3.06
Volatility (%)	2.79	2.83	2.82	2.80	2.80	2.84	2.83	2.87	2.84	2.81	2.73	2.81	2.85	2.74	2.63	2.59	2.72	2.49	1.89	2.66	2.26	2.73
Sharpe ratio	0.70	0.68	0.68	0.66	_0.74_	_0.75_	_0.72_	_0.79*_	_0.74_	0.69	0.68	_0.71_	_0.83**_	_0.93***_	_0.83**_	_0.82*_	_1.02***_	0.65	_1.04*_	0.66	0.61	0.66
Port. turnover (%)	2.16	2.25	2.23	2.22	2.26	2.24	2.24	2.26	2.24	2.23	2.41	2.25	2.28	2.31	2.25	2.35	2.30	2.27	2.61	2.37	2.54	2.25
Avg. com. weight (%)	-	3.02	1.99	1.54	3.48	2.49	2.47	3.30	2.49	1.82	6.17	2.68	3.61	7.30	3.68	7.54	7.42	11.53	49.95	6.04	19.61	2.78
Panel E: Reward-to-risk timing																						
Return (%)	3.20	3.24	3.24	3.09	3.62	3.83	3.67	4.11	3.79	3.23	3.40	3.34	4.68	4.86	4.03	4.08	5.76	3.23	3.67	3.15	2.77	3.06
Volatility (%)	2.96	3.08	3.21	3.05	3.30	3.41	3.48	3.61	3.50	3.20	2.85	3.17	3.75	3.31	2.81	2.90	3.43	2.85	2.09	2.69	2.70	2.91
Sharpe ratio	0.66	0.65	0.63	0.61	_0.73_	_0.77_	_0.71_	_0.80_	_0.74_	0.63	_0.76_	_0.67_	_0.93_	_1.10**_	_1.00*_	_0.99*_	_1.31***_	_0.70_	_1.22***_	_0.71_	0.57	0.62
Port. turnover (%)	5.40	5.56	5.54	5.50	5.54	5.43	5.43	5.38	5.41	5.52	5.65	5.53	5.42	5.42	5.33	5.32	5.32	5.63	5.65	5.80	6.28	5.58
Avg. com. weight (%)	-	3.56	3.50	1.73	6.70	5.88	5.86	7.76	6.18	3.52	7.75	4.84	9.33	17.18	8.31	12.63	21.85	7.58	44.66	6.23	3.61	0.90

Notes: This table reports the results of several out-of-sample asset allocation strategies (Panel A to Panel D) for a benchmark (stock–bond) portfolio and the benchmark portfolio complemented with one out of 21 commodity indices at a time. The results are derived from a rolling sample approach with a 60-month estimation window and cover the full sample from February 2000 to April 2017. Return and volatility denote the annualized time series mean and volatility of monthly net returns, respectively. The Sharpe ratio is calculated as annualized time series mean of net excess returns divided by the annualized time series volatility of net excess returns. 'Port. turnover' denotes the portfolio turnover, calculated as the average change in weights over all rebalancing points. 'Avg. com. weight' denotes the average weight in the commodity index over the full sample. Italic font indicates an absolute increase in the Sharpe ratio of the augmented portfolio relative to the benchmark case. *, ** and *** indicate a significant increase in the Sharpe ratio of the augmented portfolio relative to the benchmark case at the 10%, 5% and 1% level, respectively.

Hannover Center of Finance e.V. Diskussionspapiere / Discussion Papers

Hannes Rehm, Die Europäische Bankenunion – ein Element der internationalen Bankordnungspolitik, #1, 28. Januar 2015.

Rouven Wiegard, Domenic Teuber und Michael H. Breitner, Verwendung von Immobilienvermögen zur Alterssicherung am Beispiel der Umkehrhypothek, #2, 16. Februar 2015.

Rouven Wiegard, Nabileh Rahmati und Michael H. Breitner, Rentenversicherungssystem: Quo vadis?, #3, 4. März 2015.

Hannes Rehm, Die Zukunft der Bankindustrie – Bestimmungsgründe und Entwicklungslinien künftiger Geschäftsmodelle, #4, 29. April 2016.

Hannes Rehm, Die Krise der Gemeinschaft als Chance für den Finanzplatz Europa – gestalten statt verwalten, #5, 27. Februar 2017.

Tobias Fethke und Marcel Prokopczuk, Is Commodity Index Investing Profitable?, #6, 5. April 2018.

Impressum: Hannover Center of Finance e.V., Königsworther Platz 1, 30167 Hannover, Vorstand: Prof. Dr. Michael H. Breitner (Sprecher), Prof. Dr. Johann-Matthias Graf von der Schulenburg (stellv. Sprecher) und Prof. Dr. Marcel Prokopczuk.