Journal of Empirical Finance 58 (2020) 164–180

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Does sophistication of the weighting scheme enhance the performance of long-short commodity portfolios?[☆]

Hossein Rad [a,*], Rand Kwong Yew Low [a,b], Joëlle Miffre [c], Robert Faff [a]

[a] UQ Business School, The University of Queensland, Brisbane, 4072, Australia
[b] Bond Business School, Bond University, Gold Coast, 4226, Australia
[c] Audencia Business School, 8 Route de la Jonelière, 44312, Nantes, France

ARTICLE INFO

JEL classification:
G13
G14

Keywords:
Long-short portfolios
Equal weights
Optimized weights
Risk-timing weights

ABSTRACT

The commodity pricing literature advocates the design of long-short portfolios based on equal weights. Relaxing the assumption of naive diversification, this article studies the benefits of applying sophisticated weighting schemes to the construction of long-short momentum and term structure portfolios. Weighting schemes based on risk minimization and risk timing are found to dominate the naive allocation and the weighting schemes based on utility maximization. This conclusion is not challenged by concerns pertaining to transaction costs, illiquidity, data mining, sub-periods, and model parameters and robustly persists when we consider as sorting signals hedging pressure, speculative pressure and, to a lower extent, basis-momentum.

1. Introduction

There are theoretical and empirical reasons to believe that commodity futures investments command positive risk premia. The theoretical considerations relate either to the theory of storage of Kaldor (1939) where the risk premium depends on inventory levels, and thus on the slope of the futures curve, or to the hedging pressure hypothesis of Cootner (1960) where the risk premium is a function of hedgers' and speculators' net positions. These theories have been empirically validated in numerous studies[1] that advocate long positions in backwardated futures and short positions in contangoed futures.[2] While less theoretically grounded, other signals have also been shown to successfully predict commodity futures price changes, including: past performance,[3] value, volatility, open interest, skewness, or basis-momentum (Erb and Harvey, 2006; Miffre and Rallis, 2007; Hong and Yogo, 2012; Asness et al., 2013; Szymanowska et al., 2014; Miffre, 2016; Fernandez-Perez et al., 2018; Boons and Prado, 2019).

☆ The paper has benefited from discussions with J. Fan, A. Fernandez-Perez, A.-M. Fuertes, M. Prokopczuk, K. Walsh and participants at the Griffith Alternative Investments 2016 conference, Gold Coast, Australia, at the Energy and Commodity Finance 2017 conference, Oxford, UK, and at the 2017 International Accounting and Finance Doctoral Symposium, Warsaw, Poland. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author.

E-mail addresses: h.rad@business.uq.edu.au (H. Rad), r.low@business.uq.edu.au, rlow@bond.edu.au (R.K.Y. Low), jmiffre@audencia.com (J. Miffre), r.faff@business.uq.edu.au (R. Faff).

[1] Support for the theory of storage is provided in Fama and French (1987), Erb and Harvey (2006), Gorton and Rouwenhorst (2006), Symeonidis et al. (2012), Gorton et al. (2013), and Szymanowska et al. (2014). Evidence in favor of the hedging pressure hypothesis can be found in Bessembinder (1992), De Roon et al. (2000), Basu and Miffre (2013), and Kang et al. (2020).

[2] Backwardation predicts a rise in commodity futures prices driven by scarce inventories, a downward-sloping term structure of futures prices, net short hedging, or net long speculation. Conversely, contango predicts a drop in commodity futures prices driven by abundant inventories, an upward-sloping term structure, net long hedging, or net short speculation.

[3] Past performance or momentum has been shown to capture the phases of backwardation and contango as winners (losers) present backwardated (contangoed) characteristics such as positive (negative) roll yields, net short (long) hedging, net long (short) speculation, and low (high) standardized inventories (Miffre and Rallis, 2007; Gorton et al., 2013).

https://doi.org/10.1016/j.jempfin.2020.05.006
Received 12 July 2018; Received in revised form 27 January 2020; Accepted 24 May 2020
Available online 2 June 2020

In prior studies, commodities enter the portfolios with equal weights. The rationale for this naive choice comes from the fact that contrasting equity markets, there is no natural value-weighting scheme that can be easily applied to commodity portfolios; the closest equivalent (i.e., production and consumption weighting) is difficult to implement given the challenges of collecting reliable inventory data (Gorton et al., 2013; Symeonidis et al., 2012).[4]

Our key contribution is to relax the assumption of naive equal weights and to apply sophisticated weighting schemes to the design of long-short commodity portfolios. These sophisticated schemes emanate from two strands of the equity literature. The first one defines the asset allocation based on an optimization procedure (Markowitz, 1952; DeMiguel et al., 2009; Kolm et al., 2014); we consider five variants thereof based on either utility maximization or risk minimization. The second strand is based on risk timing as originated by Kirby and Ostdiek (2012) (volatility timing, reward-to-risk timing, and beta timing) and as subsequently extended in the present article (Value-at-Risk timing and conditional Value-at-Risk timing).

The idea is to apply the aforementioned weighting schemes to the construction of commodity-based long-short momentum and term structure portfolios. By doing this, the proposed optimized and risk-timing strategies not only capture the fundamentals of backwardation and contango that are key to the pricing of commodity futures, they also assign sophisticated weights to the constituents of these portfolios. We can then compare the out-of-sample performance of the proposed sophisticated momentum and term structure strategies to that of their naive equal-weight counterparts and thereby test whether such sophistication of the weighting schemes enhances performance.

Our findings suggest that the naive-weighting scheme traditionally employed in the construction of long-short commodity portfolios can be challenged by the use of sophisticated weighting schemes. Momentum and term structure portfolios based on weights that minimize or time risk present better performance than their naive counterparts based on equal weights. At the other end of the spectrum, long-short portfolios based on utility maximization generate a risk-adjusted performance that is worse than that of the naive portfolios. Following Kirby and Ostdiek (2012), we attribute the better out-of-sample performance of the risk-timing schemes to lower estimation errors and reasonable turnover and the poor performance of the utility-maximizing schemes to estimation risk, extreme long-short weights, and high turnover.

The outperformance of the momentum and term structure portfolios based on risk-minimization and risk-timing allocations persist after accounting for the risks of the strategies as captured by, for example, the three-factor model of Bakshi et al. (2019). The conclusion is not challenged by concerns pertaining to transaction costs, illiquidity, data mining, sub-periods, and model parameters. Our finding is robust to the consideration of alternative signals for the phases of backwardation and contango as proxied by the net positions of hedgers and speculators (Basu and Miffre, 2013) and also holds for the basis-momentum portfolio of Boons and Prado (2019), albeit less strongly.

Our paper contributes to the literature on portfolio construction by proposing a novel approach that designs long-short portfolios based on optimized and risk-timing weights. Designing such a long-short framework is crucial as volatility timing weights are by construction positive and optimized weights of either sign tend to be extreme in the presence of estimation errors, leading to high turnover, poor out-of-sample performance, and trivial investment practicality (Jagannathan and Ma, 2003; Michaud, 2008; Kirby and Ostdiek, 2012; Low et al., 2016). Our work also speaks to the literature on portfolio performance and, more specifically, to that on the relative merits of various weighting schemes (DeMiguel et al., 2009; Tu and Zhou, 2011; Kirby and Ostdiek, 2012; Kolm et al., 2014). By showing that risk-timing portfolios can outperform equal-weight portfolios in the context of long-short asset allocation strategies, our analysis extends to a long-short setting the conclusions that Kirby and Ostdiek (2012) drew in a long-only setting.

The remainder of the paper is structured as follows. Sections 2 and 3 present the research method and the dataset, respectively. Section 4 discusses the empirical results. Section 5 provides robustness checks and finally, Section 6 concludes.

2. Research method

2.1. Allowing for weights of either sign

Since commodity futures markets switch from backwardation to contango depending on supply and demand shocks, commodity futures strategies should freely allow for long, as well as short, weights. The simplest way to accommodate this specificity is by removing short-selling constraints from the optimization function, thereby allowing for weights of either sign. Unfortunately, this leads to extreme weights in the presence of estimation errors, poor out-of-sample performance, and trivial investment practicality (Jagannathan and Ma, 2003; Michaud, 2008; Kirby and Ostdiek, 2012). Allowing for negative weights within the risk-timing strategies of Kirby and Ostdiek (2012) is not trivial either: the approach defines portfolio weights based on various risk measures, which by definition are strictly positive.

To accommodate the need for negative weights, we adapt the portfolio construction method used for long-only portfolios as follows. First, we decipher whether a given commodity is backwardated or contangoed and consequently, whether it shall be bought or sold. Second, over the ranking period used to define the optimized weights (typically, one year of daily data), the returns of the contangoed commodities are multiplied by -1 and the returns on the backwardated commodities are left unchanged. Third, the matrix of returns thereby obtained is used to define $X = (x_1, \ldots, x_N)'$, a $N \times 1$ vector of positive weights ensuring that the conditions of full collateralization and full investment are imposed; namely, $\Sigma_{i=1}^{N} x_i = 1$. Section 2.2.1 (Section 2.2.2) details how these weights

[4] The required data on inventories is to be collected in reference to the delivery place of the underlying asset of the futures contract and is not always publicly available. Furthermore, inventories are often revised after being published, making it difficult to analyze the profitability of the signal.

are estimated for the optimized (risk-timing) portfolios. Fourth, the long-short portfolio constructed is held over the next month. The approach is rolled forward a month and at that time a new portfolio is constructed. As the weights calculated at the end of a given month are used out-of-sample to structure the portfolios over the following month, the analysis does not suffer from perfect-foresight bias. As such, it is of value to asset managers interested in designing practical investment solutions.

Two signals are used to tell apart whether a given commodity futures is backwardated or contangoed in the first step above. The first signal for the momentum portfolio is based on past performance, measured as excess return over the year preceding portfolio formation, where a positive (negative) excess return triggers a long (short) recommendation. The second signal for the term structure portfolio is based on roll yield measured at the time of portfolio formation as the differential in the log of the settlement prices of the nearby and second nearby contracts; a positive (negative) roll yield prompts a long (short) position. The motivation comes from the fact that commodities with good past performance or positive roll yields are in backwardation and thus expected to appreciate, while commodities with poor past performance or negative roll yields are in contango and thus expected to depreciate. Alternative signals based on hedging pressure, speculative pressure and basis-momentum are considered as robustness checks in Section 5.

Our naive and sophisticated momentum portfolios do not merely invest in extreme performers; rather, they consider the N assets present at the time of portfolio formation. As such, they resemble the time-series trend-following portfolios of Szakmary et al. (2010) and Moskowitz et al. (2012) that are based on equal weights. Equivalently, our naive and sophisticated term structure portfolios could be regarded as time-series term structure portfolios. Such a time-series approach ensures better diversification and allows for the consideration of a wider range of portfolios. Indeed, depending on market conditions, the portfolios formed may be long-only (if all futures are deemed backwardated), short-only (if all futures are deemed contangoed), or long-short (if futures have a mixed classification between backwardation and contango). By contrast, the cross-sectional approach typically considered in the literature (e.g., Erb and Harvey, 2006; Miffre and Rallis, 2007) consistently forces the portfolios to be long-short.

2.2. Sophisticated weighting schemes

2.2.1. Optimized weights

Three of the five optimized solutions considered maximize, at time t, investor's expected utility of wealth at time $t+1$ with respect to portfolio weights: $\max_{x_{i,t}} E_t[U(W_{t+1})]$. Without loss of generality, we normalize investor's initial wealth to 1, and therefore have

$$W_{t+1} = W_t(1 + r_{p,t+1}) = W_t(1 + \Sigma_{i=1}^{N} x_i r_{i,t+1}) = 1 + \Sigma_{i=1}^{N} x_i r_{i,t+1}, \tag{1}$$

under the constraints of positive weights $x_i \geq 0, \quad i = 1, \ldots, N$, full investment and full collateralization $\Sigma_{i=1}^{N} x_i = 1$. Portfolio weights are defined under the following utility functions:

- Power utility (PU):

$$U(W) = \frac{W^{1-\gamma} - 1}{1 - \gamma}, \gamma \neq 1 \tag{2}$$

- Negative exponential utility (NE):

$$U(W) = \frac{-e^{-\eta W}}{\eta}, \eta \geq 0 \tag{3}$$

- Power utility with disappointment aversion (DA) and 0 as the reference point relative to which gains and losses are measured (Gul, 1991):

$$U(W) = \begin{cases} \frac{W^{1-\gamma}-1}{1-\gamma} & if \quad W > 0, \\ \frac{W^{1-\gamma}-1}{1-\gamma} + (\frac{1}{A} - 1)[\frac{W^{1-\gamma}-1}{1-\gamma}] & if \quad W \leqslant 0, \end{cases} \tag{4}$$

γ is the coefficient of relative risk aversion ($\gamma = 3$), η is the coefficient of absolute risk aversion ($\eta = 3$), $A \leq 1$ is the coefficient of disappointment aversion ($A = 0.6$), where A captures the fact that investors are more sensitive to losses than to gains of equal size.

Portfolio weights are also calculated for a mean-variance (MV) portfolio:

$$\max_{X_t} (X_t' \mu - \frac{\gamma}{2} X_t' \Sigma X_t), \tag{5}$$

where X_t is the vector of weights and μ and Σ are the estimated excess returns vector and covariance matrix, respectively.[5] Lastly, we consider the minimum variance (MIN) portfolio, assuming therefore that expected returns can be ignored. Portfolio weights are obtained as follows:

$$\min_{X_t} X_t' \Sigma X_t \tag{6}$$

[5] Three reasons govern our choice of utility functions. First, originating from Modern Portfolio Theory, MV comes across as a natural weighting scheme as it is optimal under the simplifying assumptions of quadratic utility or normally distributed returns. Second, the other utility functions (PU, NE, and DA) are considered as they allow for potential departures of asset returns from normality and present features that model the real behavior of rational risk-averse investors (monotonicity, concavity, and/or loss aversion). Third, we note that these utility functions are very frequently used, not only in the equity literature (Kandel and Stambaugh, 1996; Driessen and Maenhout, 2007; DeMiguel et al., 2009; Kostakis et al., 2011) but also in the commodity literature when studying the potential utility gains obtained while adding commodities to portfolios of traditional assets (Daskalaki and Skiadopoulos, 2011; Gao and Nardari, 2018). Thus, while we do not claim that these utility functions are superior to every utility function ever proposed, we see them as natural candidates to challenge the superiority of the equally-weight schemes in the context of momentum and term structure strategies.

2.2.2. Risk-timing weights

The risk-timing schemes employed do not use any formal optimization and as such, they are often considered as feasible alternatives to the optimized solutions (the latter are known to lead to extreme weights and poorly diversified allocation). In total, we consider five risk-timing schemes; three define asset allocation based on volatility timing, beta timing, and reward-to-risk timing (Kirby and Ostdiek, 2012); the other two are based on tail risk as modeled via Value-at-Risk and conditional Value-at-Risk; to the best of our knowledge, they are novel to the literature on portfolio construction.

Volatility timing (VT) weights. The first class of risk-timing weights uses exclusively the assets' time-varying volatilities. This is done by setting all off-diagonal elements of the covariance matrix to zero.[6] Formally, portfolio weights are given by

$$x_{i,t} = \frac{1/\sigma^2_{i,t}}{\sum_{i=1}^{N} 1/\sigma^2_{i,t}}, \quad i = 1, 2, \dots, N \tag{7}$$

By introducing a tuning parameter, η, in the above equation, the volatility timing (VT) strategy is generalized to allow for more flexibility in terms of how aggressively the portfolio weights respond to changes in volatility

$$x_{i,t} = \frac{(1/\sigma^2_{i,t})^{\eta}}{\sum_{i=1}^{N} (1/\sigma^2_{i,t})^{\eta}}, \quad i = 1, 2, \dots, N, \quad \eta \geq 0 \tag{8}$$

As $\eta \longrightarrow 0$, the effect of volatilities on determining portfolio weights diminishes and the strategy moves towards the equally-weighted (EW) strategy. As $\eta \longrightarrow \infty$, the strategy assigns larger weights to assets with lower volatility.

Reward-to-risk timing (RRT) weights. Although expected returns are usually estimated with far less precision than volatilities (Merton, 1980), treating them as part of the portfolio construction process enhances the available information set which can lead to better performance. The second class of risk-timing weights therefore incorporates conditional expected returns in Eq. (8) as follows

$$x_{i,t} = \frac{(\mu^+_{i,t}/\sigma^2_{i,t})^{\eta}}{\sum_{i=1}^{N} (\mu^+_{i,t}/\sigma^2_{i,t})^{\eta}}, \quad i = 1, 2, \dots, N, \quad \eta \geq 0, \tag{9}$$

$\mu^+_{i,t} = max\{0, \mu_{i,t}\}$, where $\mu_{i,t}$ denotes average excess returns as estimated for asset i at time t. The restriction $\mu^+_{i,t} \geq 0$ is needed for the reward-to-risk timing (RRT) term structure portfolio to ensure that $x_{i,t} \geq 0, \quad \forall i, t$;[7] it is redundant for the RRT momentum portfolio for which $\mu^+_{i,t} \geq 0$ and thus $x_{i,t} \geq 0$ by default. The base case RRT strategy assumes $\eta = 1$.

Beta timing (BT) weights. The third class of risk-timing weights addresses concerns pertaining to the high degree of estimation errors in sample expected returns (Merton, 1980). This is done by estimating conditional expected returns using the conditional CAPM as follows

$$E_t(r_{i,t+1}) = \beta_{i,t} E_t(r_{M,t+1}), \quad i = 1, 2, \dots, N, \tag{10}$$

where $r_{i,t+1}$ and $r_{M,t+1}$ denote the time $t+1$ excess return of asset i and that of the market, respectively. To ensure $x_{i,t} \geq 0, \quad i = 1, 2, \dots, N$, we assign strictly positive weights to the assets whose betas have the same sign as the conditional expected return of the market. Replacing expectations with sample moments in Eq. (10) and $\mu^+_{i,t}$ by $\beta^+_{i,t}$ in Eq. (9), the beta timing (BT) weights are defined as

$$x_{i,t} = \frac{(\beta^+_{i,t}/\sigma^2_{i,t})^{\eta}}{\sum_{i=1}^{N} (\beta^+_{i,t}/\sigma^2_{i,t})^{\eta}}, \quad i = 1, 2, \dots, N, \quad \eta \geq 0, \tag{11}$$

where $\beta^+_{i,t} = |\beta_{i,t}|$ if $\beta_{i,t}$ has the same sign as $\mu_{M,t}$ and 0 otherwise (to ensure $x_{i,t} \geq 0 \quad for \quad i = 1, \dots, N$). $\beta_{i,t}$ is estimated in relation to the daily excess returns of the S&P-GSCI. As before for VT and RRT, the base case BT strategy assumes $\eta = 1$.

Value-at-Risk timing (VaRT) weights. Volatility is the sole risk influencing the VT weights. However, it suffers from the shortcomings of focusing on upside (as well as downside) volatility and of ignoring tail risk. To address these problems, we estimate Value-at-Risk timing weights which employ Value-at-Risk (VaR) in lieu of volatility in Eq. (8)

$$x_{i,t} = \frac{(1/v^2_{i,t})^{\eta}}{\sum_{i=1}^{N} (1/v^2_{i,t})^{\eta}}, \quad i = 1, 2, \dots, N, \quad \eta \geq 0, \tag{12}$$

where v denotes the α level empirical VaR. We set $\alpha = 0.95$ which is among the most-frequently used values for VaR in practice (Low et al., 2013). Again, the base case Value-at-Risk timing (VaRT) strategy assumes $\eta = 1$.

Conditional Value-at-Risk timing (CVaRT) weights. VaR suffers from two known shortcomings; that of ignoring the magnitude of losses greater than VaR and that of not being a coherent risk measure (Artzner et al., 1999), resulting in the underestimation of extreme

[6] The assumption of zero return correlations reduces estimation error and turnover, resulting in potentially enhanced net performance.

[7] To ensure $x_{i,t} \geq 0 \quad for \quad i = 1, 2, \dots, N$ in Eq. (9), the RRT term structure strategy assigns zero weights to the contangoed commodities with negative roll yields and negative past performance (in that case, $\mu^+_{i,t} = 0$ and $x_{i,t} = 0$) and strictly positive weights to the contangoed commodities with negative roll yields and positive past performance (in that case, $\mu^+_{i,t} = \mu_{i,t}$ and $x_{i,t} > 0$). As contangoed assets with negative roll yields shall be sold, these positive weights are subsequently multiplied by -1 (as detailed in Section 2.1).

losses. Conditional Value-at-Risk (CVaR), also known as expected shortfall, circumvents these issues by reporting the expected value of losses below the corresponding VaR threshold. Conditional Value-at-Risk (CVaR) is subadditive and convex, thus it is a coherent risk measure. It is also easier to estimate (Rockafellar and Uryasev, 2002). The conditional Value-at-Risk timing (CVaRT) weights are calculated by substituting VaR with CVaR in Eq. (12)

$$x_{i,t} = \frac{(1/c_{i,t}^2)^\eta}{\sum_{i=1}^N (1/c_{i,t}^2)^\eta}, \quad i = 1, 2, \ldots, N, \quad \eta \geq 0, \tag{13}$$

where c denotes the α level empirical CVaR and $\alpha = 0.95$. The base case CVaRT strategy assumes $\eta = 1$.

2.3. Risk and performance evaluation

We appraise the riskiness of each portfolio by way of the following measures of risk: annualized total volatility, annualized downside volatility (using 0% as threshold), skewness, excess kurtosis, Jarque–Bera normality test, 95% VaR, 95% CVaR and maximum drawdown. We further measure performance using annualized mean excess return, Sharpe ratio (defined as mean over total volatility), Sortino ratio (defined as mean over downside volatility) and Omega ratio (defined as the probability of gains divided by the probability of losses using 0% as threshold). Further evidence of abnormal performance is brought forward through OLS regressions of the excess returns of a sophisticated strategy onto i) the excess returns of the corresponding EW strategy or ii) the three-factor benchmark of Bakshi et al. (2019). The idea is to test whether the strategies implied by the sophisticated weighting schemes merely capture the risk premia known to be present in commodity futures markets or whether they are capable of generating positive and significant alphas.

3. Data and sampling

The dataset, from Datastream International, consists of daily prices for 40 commodity futures from January 1979 to April 2016. The constituents of the dataset comprise 10 metal commodity futures (aluminum, copper, gold, lead, nickel, palladium, platinum, silver, tin, zinc), 4 livestock (feeder cattle, frozen pork bellies, lean hogs, live cattle), 9 energy contracts (Brent crude oil, coal, electricity, gas oil, gasoline, heating oil, light sweet crude oil, natural gas, unleaded gasoline), 14 agricultural commodities (cocoa, coffee, corn, cotton, frozen concentrated orange juice, oats, rough rice, soybeans, soybean meal, soybean oil, sugar number 11, sugar number 14, wheat, white wheat), alongside with the futures on lumber, milk and western plywood. Our study assumes that investors hold fully-collateralized positions in these contracts. Excess returns are measured as the percent changes in daily settlement prices using either nearest or second nearest contracts; the second nearest contracts are used in months when the nearest contracts mature. This approach addresses potential liquidity issues.

The article compares the performance of the naive and sophisticated momentum and term structure portfolios to benchmarks that are now standard in the commodity pricing literature (Erb and Harvey, 2006; Miffre and Rallis, 2007; Szymanowska et al., 2014; Bakshi et al., 2019). These include: (i) AVG, a long-only equally-weighted and monthly-rebalanced portfolio of all commodity futures, (ii) the S&P-GSCI, (iii) CARRY, an equally-weighted monthly-rebalanced portfolio that is long the 5 commodity futures with highest roll yields and short the 5 commodities with lowest roll yields, (iv) MOM, an equally-weighted monthly-rebalanced portfolio that is long the 5 commodity futures with highest mean excess returns over the past 6 months and short the 5 commodity futures with lowest mean excess returns over the past 6 months and (v) LIQ, a long-short portfolio that is long the 5 commodity futures with lowest "Amivest" measure of liquidity and short the 5 commodity futures with highest "Amivest" measure of liquidity (Amihud et al., 1997; Marshall et al., 2012).[8] Appendix presents summary statistics for these portfolios. Aligned with the literature (Miffre, 2016), the long-short portfolios generate better performance than the long-only benchmarks.

4. Main empirical results

4.1. Momentum strategies

Can the optimized and risk-timing strategies proposed in Section 2 be used to enhance the performance of the standard (EW) momentum strategy? Table 1, Panel A answers the question by looking at various performance measures. With Sharpe and Sortino ratios at 0.74 and 1.08 respectively, the naive EW momentum strategy ranks 5th out of the 11 momentum strategies considered. MIN and VT are shown to present the highest Sharpe ratios (0.97 and 0.95, respectively), the highest Sortino ratios (1.45 and 1.50, respectively) and the highest Omega ratios (2.12 and 2.06, respectively). VaRT and CVaRT rank closely thereafter with e.g., Sharpe ratios at 0.87 and Sortino ratios exceeding 1.3. At the other end of the spectrum, the momentum strategies that maximize expected utility rank worst with e.g., Sharpe ratios around 0.3.

The outperformance of MIN, VT, VaRT and, CVaRT relative to EW does not come from a better gross performance (the mean excess returns obtained are often statistically the same). Rather the higher Sharpe, Sortino and Omega ratios are driven by their more appealing risk profiles; namely, their lower volatility, lower downside volatility, lower maximum drawdown, higher VaR, or higher CVaR. The strategies based on expected utility maximization suffer from risk characteristics that are unattractive such as

[8] Amivest measure is calculated as the average over the 2 months preceding portfolio formation of the ratio of daily $volume (or volume multiplied by settlement price) to daily absolute return.

Table 1
Risk and performance of momentum portfolios under various weighting schemes. Panel A presents summary statistics for the excess returns of long-short momentum portfolios under various weighting schemes. Panels B and C report parameter estimates, Newey–West *t*-statistics, and adjusted R^2 obtained from regressions of the excess returns of a given naive or sophisticated strategy on the excess returns of the relevant risk factors. PU, NE, DA, MV, and MIN stand for power utility, negative exponential utility, power utility with disappointment aversion, mean variance utility, and minimum variance, respectively. VT, RRT, BT, VaRT and CVaRT stand for volatility timing, reward-to-risk timing, beta-timing, Value-at-Risk timing, and Conditional Value-at-Risk timing, respectively. Mean, volatility, downside volatility and abnormal returns (α) have been annualized, *t*-statistics for the null hypothesis that a given sophisticated strategy generates the same mean excess return as its equal-weight counterpart are reported under the heading "ew tstat". Bold fonts denote significance at the 5% level or better. AVG is a long-only equally-weighted monthly-rebalanced portfolio of all commodity futures, CARRY and MOM are equally-weighted monthly-rebalanced long-short portfolios based on roll yields and past performance, respectively. The analysis spans the sample January 1980 to April 2016.

	Equal	*Optimized weights*					*Risk-timing weights*				
	Weights	PU	NE	DA	MV	MIN	VT	RRT	BT	VaRT	CVaRT
Panel A: Summary statistics											
Mean	**0.0619**	0.0749	0.0739	0.0749	0.0779	**0.0588**	**0.0612**	**0.0919**	**0.0813**	**0.0625**	**0.0620**
tstat	(3.80)	(1.06)	(1.22)	(1.06)	(1.35)	(4.23)	(4.04)	(3.55)	(2.39)	(4.06)	(4.08)
ew tstat		(−0.24)	(−0.26)	(−0.24)	(−0.30)	(0.38)	(0.08)	**(−2.33)**	(−0.79)	(−0.13)	(−0.01)
Volatility	0.08	0.35	0.31	0.35	0.35	0.06	0.06	0.13	0.19	0.07	0.07
Downside volatility	0.06	0.25	0.23	0.25	0.25	0.04	0.04	0.10	0.13	0.05	0.05
Skewness	−0.22	**−0.35**	**−0.46**	**−0.35**	**−0.34**	−0.07	0.12	**−0.57**	−0.17	−0.03	0.01
Ex. kurtosis	**1.64**	**3.40**	**4.00**	**3.40**	**3.38**	**2.37**	**2.13**	**4.82**	**2.75**	**2.22**	**2.07**
JB p.value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VaR (95%)	−0.04	−0.16	−0.13	−0.16	−0.16	−0.02	−0.03	−0.05	−0.08	−0.03	−0.03
CVaR (95%)	−0.05	−0.24	−0.21	−0.24	−0.23	−0.03	−0.04	−0.08	−0.12	−0.04	−0.04
Max drawdown	0.20	0.86	0.77	0.86	0.86	0.16	0.17	0.40	0.50	0.18	0.18
Sharpe ratio	0.74	0.21	0.24	0.21	0.22	0.97	0.95	0.69	0.44	0.87	0.87
Sortino ratio	1.08	0.29	0.32	0.29	0.31	1.45	1.50	0.92	0.63	1.31	1.34
Omega ratio	1.76	1.18	1.21	1.18	1.19	2.12	2.06	1.73	1.40	1.95	1.94
Panel B: Regression analysis relative to the equal-weight benchmark											
α		−0.0560	−0.0490	−0.0560	−0.0524	**0.0229**	**0.0236**	0.0099	−0.0083	**0.0126**	**0.0121**
		(−1.09)	(−1.11)	(−1.09)	(−1.03)	(2.98)	(3.26)	(0.84)	(−0.36)	(2.98)	(2.91)
β		**2.11**	**1.99**	**2.11**	**2.10**	**0.58**	**0.61**	**1.32**	**1.45**	**0.81**	**0.81**
		(8.45)	(8.99)	(8.45)	(8.48)	(11.96)	(15.66)	(19.01)	(14.22)	(33.91)	(35.60)
Adj. R^2		0.25	0.28	0.25	0.25	0.65	0.61	0.70	0.42	0.89	0.89
Panel C: Regression analysis relative to the (Bakshi et al., 2019) benchmark											
α	**0.0424**	−0.0219	−0.0135	−0.0219	−0.0182	**0.0440**	**0.0530**	**0.0535**	0.0412	**0.0489**	**0.0484**
	(4.26)	(−0.40)	(−0.28)	(−0.40)	(−0.33)	(5.39)	(5.41)	(3.34)	(1.47)	(5.31)	(5.30)
β_{AVG}	0.00	**0.39**	**0.35**	**0.39**	**0.38**	0.01	0.01	**0.20**	0.31	0.02	0.01
	(0.01)	(2.15)	(2.19)	(2.15)	(2.12)	(0.17)	(0.14)	(2.05)	(1.95)	(0.25)	(0.16)
β_{CARRY}	−0.01	0.28	0.25	0.28	0.28	**0.05**	−0.02	0.02	0.06	−0.02	−0.02
	(−0.22)	(1.35)	(1.33)	(1.35)	(1.33)	(2.17)	(−0.70)	(0.20)	(0.57)	(−0.57)	(−0.53)
β_{MOM}	**0.39**	**1.39**	**1.25**	**1.39**	**1.38**	**0.21**	**0.18**	**0.65**	**0.60**	**0.29**	**0.28**
	(13.67)	(9.14)	(8.97)	(9.15)	(9.16)	(8.71)	(6.09)	(10.73)	(9.77)	(9.99)	(10.14)
Adj. R^2	0.35	0.31	0.33	0.31	0.31	0.24	0.12	0.46	0.23	0.25	0.25

annualized total and downside volatilities exceeding 25%, negative skewness levels, or maximum drawdowns of more than 75%. These high levels of risk are driven by the extreme portfolio weights (i.e., poorly diversified asset allocation) that are often the outputs of optimizers (Jagannathan and Ma, 2003; Michaud, 2008; Kirby and Ostdiek, 2012).[9]

Table 1, Panel B reports parameter estimates from regressions of the excess returns of the sophisticated momentum strategies onto the excess returns of the naive EW momentum strategy. Table 1, Panel C reports similar parameter estimates but this time we treat as independent variables the excess returns of the three-factor model of Bakshi et al. (2019); namely, AVG, CARRY, and MOM. Confirming the results of Panel A, MIN, VT, VaRT, and CVaRT stand out as better strategies than EW; irrespective of the pricing model considered, these strategies generate positive and statistically significant alphas that range from 1.20% to 5.30% a year. The strategies based on utility maximization again perform poorly, destroying investors wealth relative to the benchmarks considered by an average annualized alpha of 3.62% across panels.[10]

The time-series EW momentum portfolio (that considers all assets present in the cross section) generates a Sharpe ratio at 0.74 (Table 1, Panel A) which substantially exceeds that of the cross-sectional EW momentum portfolio (that merely shortlists the 10 assets with the most extreme characteristics; Sharpe ratio of 0.40, Appendix). A comparison of the risk profiles of the two portfolios highlights the benefits of diversification obtained from holding all N assets present at the time of portfolio formation, as opposed to merely the 10 most extreme performers. Likewise, the time-series EW momentum portfolio generates a positive alpha equal to 4.24% (t-statistic of 4.26) a year relative to the three-factor benchmark of Bakshi et al. (2019). These results highlight the superiority of the time-series, over cross-sectional, approach to portfolio construction.

[9] Detailed asset allocations for all strategies are available upon request from the authors.

[10] Unlike AVG and CARRY, the MOM factor strongly and uniformly explains the excess returns of all strategies. This is not surprising as the portfolios studied in Table 1 sort commodities into backwardated and contangoed portfolios based on past performance. The adjusted R^2, however, averages 29%, suggesting that a large proportion of excess returns remains unexplained.

Table 2
Risk and performance of term structure portfolios under various weighting schemes. Panel A presents summary statistics for the excess returns of long-short term structure portfolios under various weighting schemes. Panels B and C report parameter estimates, Newey–West t-statistics and adjusted R^2 obtained from regressions of the excess returns of a given naive or sophisticated strategy on the excess returns of the relevant risk factors. PU, NE, DA, MV, and MIN stand for power utility, negative exponential utility, power utility with disappointment aversion, mean variance utility, and minimum variance, respectively. VT, RRT, BT, VaRT and CVaRT stand for volatility timing, reward-to-risk timing, beta-timing, Value-at-Risk timing, and Conditional Value-at-Risk timing, respectively. Mean, volatility, downside volatility and abnormal returns (α) have been annualized, t-statistics for the null hypothesis that a given sophisticated strategy generates the same mean excess return as its equal-weight counterpart are reported under the heading "ew tstat". Bold fonts denote significance at the 5% level or better. AVG is a long-only equally-weighted monthly-rebalanced portfolio of all commodity futures, CARRY and MOM are equally-weighted monthly-rebalanced long-short portfolios based on roll yields and past performance, respectively. The analysis spans the sample January 1980 to April 2016.

	Equal	*Optimized weights*					*Risk-timing weights*				
	Weights	PU	NE	DA	MV	MIN	VT	RRT	BT	VaRT	CVaRT
Panel A: Summary statistics											
Mean	**0.0479**	0.1041	0.0991	0.1041	0.1052	**0.0409**	**0.0279**	**0.1149**	**0.1183**	**0.0387**	**0.0384**
tstat	(2.65)	(0.81)	(0.90)	(0.81)	(1.79)	(3.52)	(2.06)	(3.58)	(2.94)	(2.46)	(2.46)
ew tstat		(−1.04)	(−1.08)	(−1.03)	(−1.06)	(0.83)	**(2.64)**	**(−3.21)**	**(−2.29)**	**(2.10)**	**(2.26)**
Volatility	0.08	0.34	0.30	0.34	0.33	0.06	0.07	0.14	0.19	0.07	0.07
Downside volatility	0.05	0.23	0.21	0.23	0.23	0.03	0.04	0.08	0.12	0.04	0.04
Skewness	**0.37**	−0.15	−0.19	−0.16	−0.15	**0.35**	**0.45**	0.22	0.21	**0.50**	**0.51**
Ex. kurtosis	**3.30**	**3.58**	**3.84**	**3.58**	**3.57**	**1.76**	**3.61**	**1.67**	**3.94**	**3.24**	**3.39**
JB p.value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VaR (95%)	−0.03	−0.14	−0.12	−0.14	−0.14	−0.02	−0.03	−0.06	−0.07	−0.03	−0.03
CVaR (95%)	−0.05	−0.21	−0.19	−0.21	−0.21	−0.03	−0.04	−0.08	−0.11	−0.04	−0.04
Max drawdown	0.22	0.86	0.78	0.86	0.86	0.13	0.17	0.30	0.57	0.18	0.18
Sharpe ratio	0.58	0.31	0.33	0.31	0.31	0.73	0.43	0.84	0.62	0.56	0.55
Sortino ratio	0.91	0.45	0.48	0.45	0.46	1.24	0.67	1.40	0.97	0.91	0.90
Omega ratio	1.57	1.28	1.31	1.28	1.28	1.76	1.41	1.91	1.65	1.55	1.54
Panel B: Regression analysis relative to the equal-weight benchmark											
α		0.0617	0.0547	0.0616	0.0630	**0.0150**	−0.0035	**0.0818**	**0.0875**	0.0003	−0.0005
		(1.10)	(1.12)	(1.10)	(1.13)	(3.07)	(−0.66)	(3.48)	(2.92)	(0.09)	(−0.14)
β		**0.89**	**0.93**	**0.89**	**0.88**	**0.54**	**0.66**	**0.69**	**0.64**	**0.80**	**0.81**
		(2.92)	(3.42)	(2.92)	(2.92)	(12.59)	(18.16)	(4.71)	(2.85)	(44.30)	(44.03)
Adj. R^2		0.05	0.07	0.05	0.05	0.64	0.70	0.18	0.08	0.91	0.92
Panel C: Regression analysis relative to the (Bakshi et al., 2019) benchmark											
α	**0.0306**	0.0058	0.0087	0.0058	0.0075	**0.0265**	**0.0214**	**0.0666**	**0.0605**	**0.0263**	**0.0263**
	(3.65)	(0.11)	(0.19)	(0.11)	(0.15)	(4.19)	(2.52)	(3.47)	(2.08)	(3.22)	(3.31)
β_{AVG}	**−0.42**	0.11	0.05	0.11	0.10	**−0.17**	**−0.31**	−0.08	0.26	**−0.35**	**−0.36**
	(−9.84)	(0.59)	(0.32)	(0.59)	(0.57)	(−6.20)	(−8.33)	(−0.78)	(1.60)	(−9.57)	(−10.21)
β_{CARRY}	**0.40**	**0.62**	**0.61**	**0.62**	**0.62**	**0.28**	**0.21**	**0.38**	**0.56**	**0.30**	**0.30**
	(14.46)	(3.03)	(3.36)	(3.03)	(3.02)	(9.54)	(5.09)	(4.62)	(5.12)	(10.26)	(10.81)
β_{MOM}	−0.03	**1.08**	**0.95**	**1.08**	**1.07**	−0.02	−0.04	**0.47**	**0.33**	−0.02	−0.02
	(−1.05)	(8.11)	(8.19)	(8.11)	(8.12)	(−1.12)	(−1.42)	(10.88)	(4.61)	(−0.68)	(−0.96)
Adj. R^2	0.68	0.26	0.28	0.26	0.26	0.46	0.46	0.38	0.23	0.62	0.63

Fig. 1 presents the evolution in the aggregated long and short weights for each of the 11 strategies considered. The plots indicate strong similarities in the long-short splits across weighting schemes. For example, the correlations between any two pairs of aggregated long weights across strategies range from 40% (between VT and BT) to 1 (between NE and DA) with an average of 69%. The darker (lighter) line on the right-hand side of each plot presents the future value of a $1 investment into the momentum (AVG) portfolio. The plots provide a graphical endorsement of the MIN, VT, VaRT, and CVaRT strategies in terms of both reducing volatility and enhancing risk-adjusted returns. Interestingly, when commodity markets enter a correction (as portrayed by a drop in the value of the AVG investment), the MIN, VT, VaRT, and CVaRT portfolios seem to post interesting returns, suggesting that these strategies can partially hedge commodity risk.

Fig. 1 also shows that PU, NE, DA, MV, and BT are the strategies that are the most sensitive to market movements; they often consist of either almost all long or almost all short positions. These near 100% long or short weighting schemes imply that there shall be near perfect synchronicity in the phases of backwardation and contango across commodity markets; this does not occur in reality as at time t the supply and demand conditions for, e.g., corn are unlikely to perfectly match those for, say, palladium. The observed extreme weights of the PU, NE, DA, MV, and BT strategies, and their lesser ability to time the specificities of each market, could explain why these strategies performed worse than the competing weighting schemes in Table 1.

We conclude that equal-weighting of the constituents of the momentum portfolio is suboptimal; enhanced risk-adjusted performance of the momentum strategy can be obtained by using weights based on risk minimization (MIN) or risk timing (VT, VaRT, and CVaRT).

4.2. Term structure-based strategies

Are our results driven by the use of past performance as signal for allocation? To answer this question, we use roll yield, in place of excess return, as sorting criterion for the backwardated and contangoed portfolios. Table 2, Panel A presents summary statistics

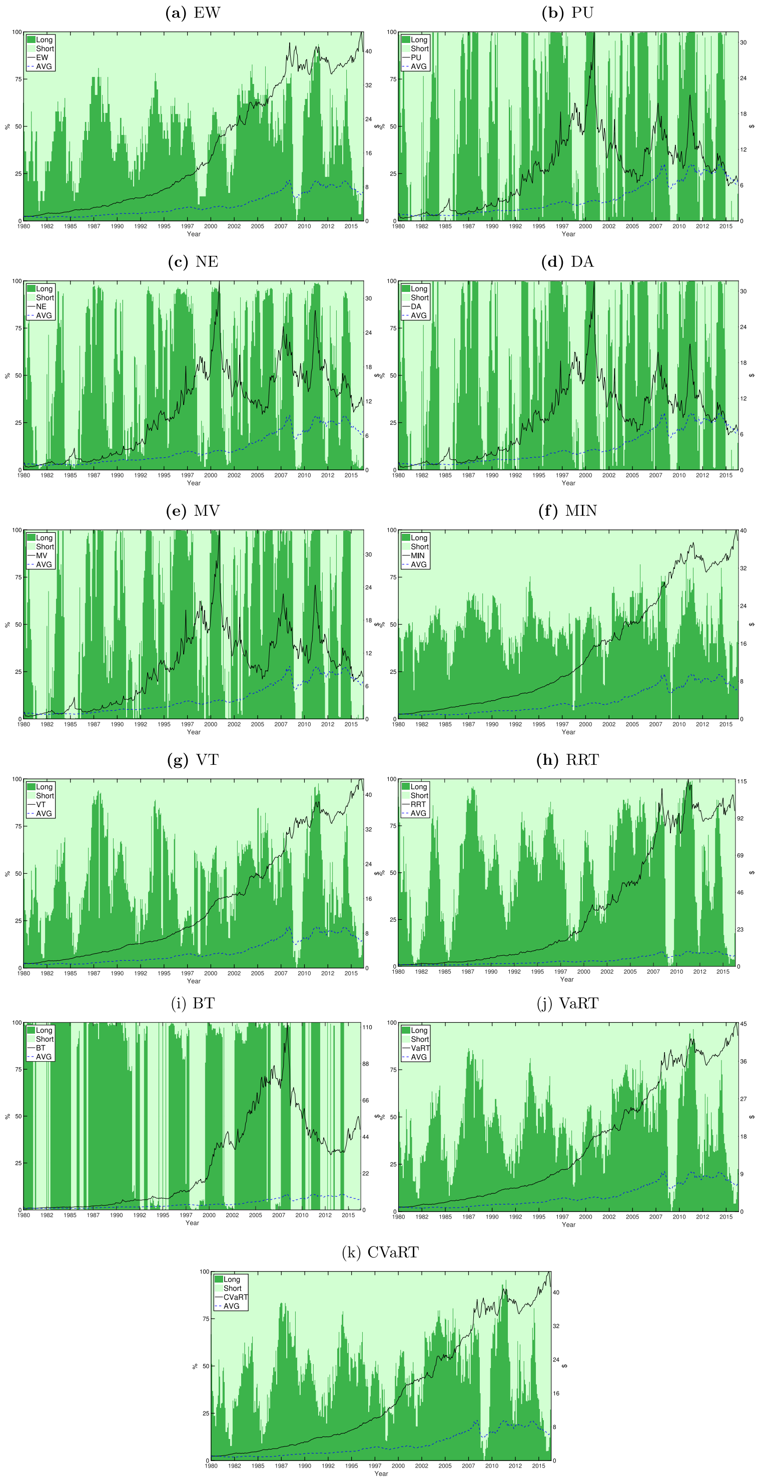

Fig. 1. Aggregated long and short momentum allocations. This figure shows the evolution of the aggregated long and short positions of the momentum strategies. The dark (light) area on each graph represents the sum of the long (short) weights. The dark (light) line represents the future value of $1 invested in the momentum (AVG) portfolio, where AVG is a long-only equally-weighted monthly-rebalanced portfolio of all commodity futures. EW stands for equal weights. PU, NE, DA, MV, and MIN stand for power utility, negative exponential utility, power utility with disappointment aversion, mean variance, and minimum variance, respectively. VT, RRT, BT, VaRT, and CVaRT stand for volatility timing, reward-to-risk timing, beta-timing, Value-at-Risk timing, and Conditional Value-at-Risk timing, respectively.

of performance for both the naive EW term structure portfolio and variants thereof based on optimized and risk-timing weights. Following the same format as Table 1, Panels B and C attempt to explain the excess returns of the sophisticated term structure strategies using either the excess returns of the naive EW term structure strategy in Panel B or the three-factor benchmark of Bakshi et al. (2019) in Panel C.

Three sophisticated term structure strategies, those based on MIN, RRT, and BT weights, stand out as outperforming the standard EW term structure portfolio. This is evidenced by Sharpe, Sortino, and Omega ratios that are higher for these sophisticated strategies and lower for the naive portfolio and by positive and statistically significant alphas in both Panels B and C (the annualized alphas range from 1.50% (MIN) to 8.75% (BT)). The outperformance of MIN, RRT, and BT relative to EW is driven either by notably higher mean excess returns (RRT and BT) or by substantially more appealing risk measures (MIN).

At the other end of the spectrum, the term structure portfolios based on utility maximization (namely, PU, NE, DA, and MV) stand out again as being the least profitable; they present Sharpe ratios that merely equal 0.32 on average (versus 0.58 for EW) and alphas that are statistically insignificant in both panels B and C. As expected (Jagannathan and Ma, 2003; Michaud, 2008; Kirby and Ostdiek, 2012; Low et al., 2016), a detailed non-reported analysis of these asset allocations indicates extreme weights and poor diversification.

Fig. 2 plots the evolution over time of the aggregated long and short allocations of the various term structure portfolios, alongside the future value of a $1 investment in the AVG and term structure portfolios. Like for momentum before, we note a tendency for the aggregated long weights to move in tandem across weighting schemes; and likewise for the aggregated short weights. The figure also confirms conclusions drawn previously. For example, panel (f) shows that MIN is a strategy with a positive and smooth performance, while panels (h) and (i) highlight that the superiority of the RRT and BT strategies in terms of performance comes at the price of relatively high levels of risk.

A comparison of the risk and performance measures obtained in Tables 1 and 2 shows that the profitability of a strategy depends more on the weighting scheme chosen than on the criterion used to sort the cross section into backwardation and contango. Irrespective of the signal employed, weighting schemes that either minimize or time risk are indeed found to beat the equal-weight benchmarks. On the other hand, weighting schemes based on utility maximization lead to poorly diversified portfolios with meagre risk-adjusted performance compared to EW.

5. Robustness checks

The conclusion thus far is that the risk-minimization and risk-timing schemes when applied to the popular EW momentum and term structure strategies generate enhanced risk-adjusted performance. This section tests the robustness of this conclusion to considerations pertaining to the choice of sorting signal, transaction costs, lack of liquidity, data mining, model parameters, and sub-periods.

5.1. Alternative sorting signals

Up until now, the article mimics the phases of backwardation and contango present in commodity futures markets via past performance or the slope of the term structure of commodity futures prices. According to the hedging pressure hypothesis of Cootner (1960) and Hirshleifer (1990), backwardation is also characterized by net short hedging and net long speculation and contango by net long hedging and net short speculation. Accordingly, we could model the risk premium of commodity futures contracts by taking positions that are opposite to those of hedgers or aligned with those of speculators.

Following Basu and Miffre (2013), we define the hedging pressure signal at each portfolio time as the standardized net position of hedgers averaged over the past 52 weeks. Formally,

$$HP_{i,t} = \frac{1}{52}\sum_{w=1}^{52}\frac{HS_{i,w} - HL_{i,w}}{HS_{i,w} + HL_{i,w}}, \tag{14}$$

where $HS_{i,w}$ and $HL_{i,w}$ are the weekly short (S) and long (L) positions of large commercial traders (also known as hedgers, H) on commodity i, respectively, as reported by the Commodity Futures Trading Commission (CFTC) in its Futures-Only Legacy Commitments of Traders report. Likewise, we define the speculative pressure signal as

$$SP_{i,t} = \frac{1}{52}\sum_{w=1}^{52}\frac{SL_{i,w} - SS_{i,w}}{SL_{i,w} + SS_{i,w}}, \tag{15}$$

where $SL_{i,w}$ and $SS_{i,w}$ are the weekly long (L) and short (S) positions of large non-commercial traders (also known as speculators, S). At each month end, we buy the presumably backwardated contracts with positive $HP_{i,t}$ or $SP_{i,t}$ signals, sell the presumably contangoed contracts with negative $HP_{i,t}$ or $SP_{i,t}$ signals, adopt the various weighting schemes detailed in Section 2 and hold the resulting naive and sophisticated portfolios on a fully-collateralized basis for a month.

Table 3 Panel A (Panel B) presents summary statistics for the performance of the hedging (speculative) pressure portfolios under the different weighting schemes. In line with our main findings, we note significant improvements in performance when moving from the standard equal weights to the sophisticated weighting schemes based on MIN and RRT. For example, the Sharpe ratio of the hedging pressure strategy increases from 0.50 for EW to 0.74 for MIN and to 0.67 for RRT; likewise for the speculative pressure portfolios. Similar increases in risk-adjusted performance are observed when analyzing the Sortino ratios, the Omega ratios or the

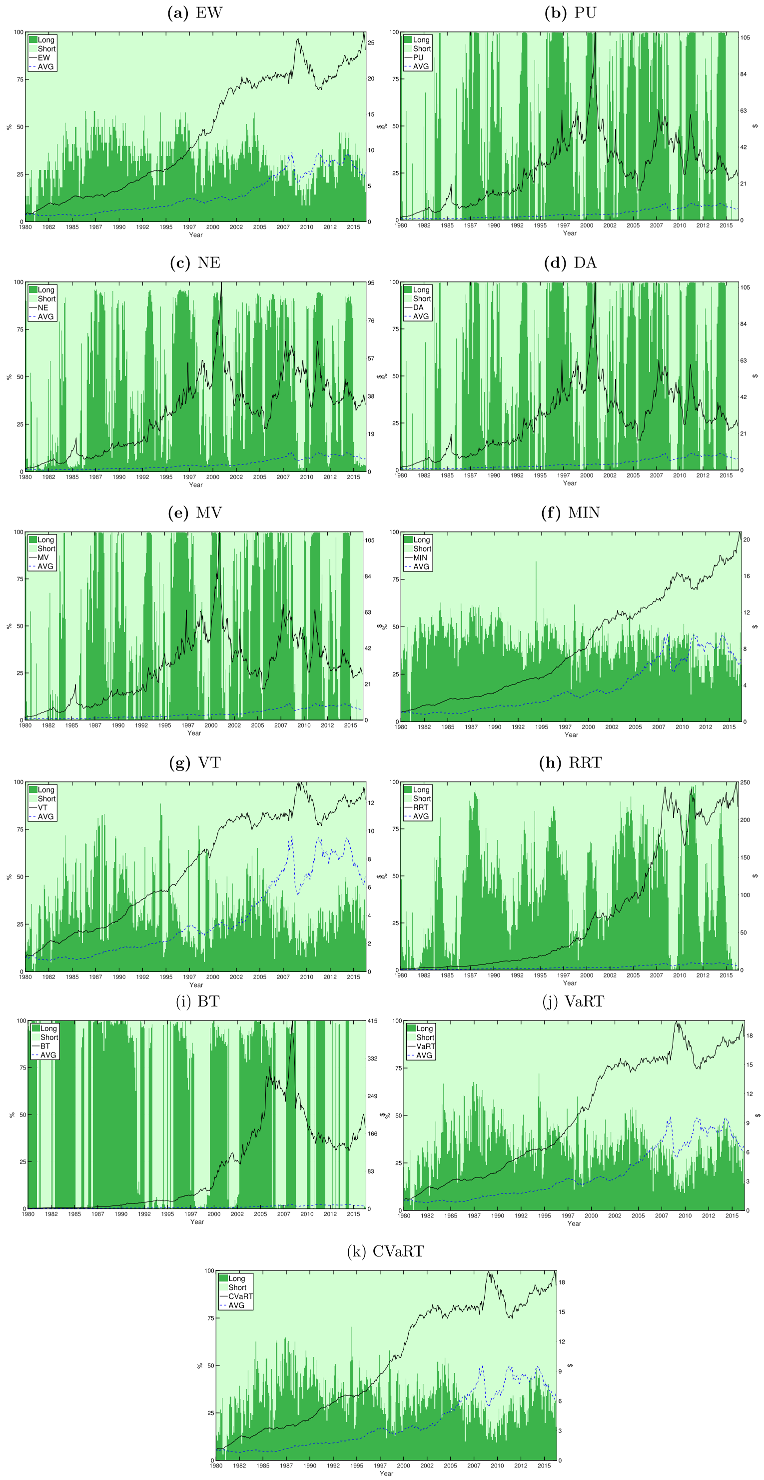

Fig. 2. Aggregated long and short term structure allocations. This figure shows the evolution of the aggregated long and short positions of the term structure strategies. The dark (light) area on each graph represents the sum of the long (short) weights. The dark (light) line represents the future value of $1 invested in the term structure (AVG) portfolio, where AVG is a long-only equally-weighted monthly-rebalanced portfolio of all commodity futures. EW stands for equal weights. PU, NE, DA, MV, and MIN stand for power utility, negative exponential utility, power utility with disappointment aversion, mean variance, and minimum variance, respectively. VT, RRT, BT, VaRT, and CVaRT stand for volatility timing, reward-to-risk timing, beta-timing, Value-at-Risk timing, and Conditional Value-at-Risk timing, respectively.

Table 3
Robustness of the results to the choice of sorting signal. The table analyzes the performance of long-short portfolios based on hedging pressure (Panel A), speculative pressure (Panel B) and basis-momentum (Panel C). α, β, Newey–West t-statistics, and adjusted R^2 are obtained from regressions of the excess returns of a given sophisticated strategy on the excess returns of the naive equal weight strategy. PU, NE, DA, MV, and MIN stand for power utility, negative exponential utility, power utility with disappointment aversion, mean variance utility, and minimum variance, respectively. VT, RRT, BT, VaRT and CVaRT stand for volatility timing, reward-to-risk timing, beta-timing, Value-at-Risk timing, and Conditional Value-at-Risk timing, respectively. Mean, volatility and abnormal returns (α) have been annualized, t-statistics for the null hypothesis that a given sophisticated strategy generates the same mean excess return as its equal-weight counterpart are reported under the heading "ew tstat". Bold fonts denote significance at the 5% level or better. The analysis spans the sample October 1993 to April 2016.

	Equal	*Optimized weights*					*Risk-timing weights*				
	Weights	PU	NE	DA	MV	MIN	VT	RRT	BT	VaRT	CVaRT
Panel A: Hedging pressure											
Mean	**0.0503**	0.1082	0.1005	0.1084	0.1098	**0.0391**	**0.0397**	**0.1082**	0.0730	**0.0457**	**0.0443**
tstat	(2.03)	(0.80)	(0.91)	(0.80)	(1.74)	(2.63)	(2.04)	(2.93)	(1.64)	(2.07)	(2.03)
ew tstat		(−0.84)	(−0.83)	(−0.84)	(−0.86)	(0.76)	(1.00)	**(−2.24)**	(−0.64)	(0.81)	(1.07)
Volatility	0.10	0.35	0.31	0.35	0.35	0.05	0.08	0.16	0.19	0.09	0.09
Sharpe ratio	0.50	0.31	0.32	0.31	0.32	0.74	0.49	0.67	0.38	0.51	0.50
Sortino ratio	0.70	0.47	0.49	0.47	0.48	1.35	0.80	1.05	0.56	0.76	0.75
Omega ratio	1.50	1.28	1.29	1.28	1.28	1.83	1.51	1.69	1.34	1.53	1.51
α		0.0542	0.0462	0.0544	0.0560	**0.0191**	0.0050	**0.0585**	0.0299	0.0026	0.0015
		(0.92)	(0.92)	(0.92)	(0.96)	(2.00)	(0.49)	(2.83)	(0.76)	(0.47)	(0.27)
β		**1.16**	**1.15**	**1.16**	**1.16**	**0.40**	**0.69**	**1.02**	**0.90**	**0.86**	**0.85**
		(4.45)	(5.05)	(4.45)	(4.49)	(8.18)	(15.81)	(10.66)	(4.96)	(31.31)	(35.22)
Adj. R^2		0.11	0.14	0.11	0.11	0.58	0.76	0.41	0.23	0.94	0.94
Panel B: Speculative pressure											
Mean	0.0409	0.0691	0.0626	0.0692	0.0711	**0.0492**	0.0425	**0.1015**	**0.0992**	0.0438	0.0400
tstat	(1.65)	(0.62)	(0.71)	(0.62)	(1.09)	(2.48)	(1.78)	(2.85)	(2.05)	(1.81)	(1.70)
ew tstat		(−0.41)	(−0.36)	(−0.41)	(−0.44)	(−0.58)	(−0.15)	**(−2.26)**	(−1.37)	(−0.53)	(0.16)
Volatility	0.10	0.35	0.31	0.35	0.34	0.06	0.09	0.17	0.22	0.09	0.09
Sharpe ratio	0.40	0.20	0.20	0.20	0.21	0.84	0.49	0.61	0.46	0.46	0.43
Sortino ratio	0.52	0.29	0.29	0.29	0.30	1.41	0.66	0.88	0.72	0.60	0.56
Omega ratio	1.40	1.17	1.18	1.17	1.18	1.97	1.50	1.62	1.42	1.48	1.43
α		0.0213	0.0159	0.0213	0.0232	**0.0311**	0.0121	**0.0592**	0.0659	0.0074	0.0038
		(0.34)	(0.30)	(0.34)	(0.38)	(3.25)	(1.23)	(2.73)	(1.63)	(1.26)	(0.73)
β		**1.17**	**1.14**	**1.17**	**1.17**	**0.44**	**0.74**	**1.03**	**0.81**	**0.89**	**0.88**
		(4.24)	(4.72)	(4.24)	(4.25)	(9.80)	(15.64)	(9.92)	(4.06)	(30.96)	(34.21)
Adj. R^2		0.12	0.14	0.12	0.12	0.60	0.78	0.41	0.14	0.94	0.94
Panel C: Basis momentum											
Mean	**0.0492**	0.0964	0.0915	0.0963	**0.0977**	**0.0348**	**0.0348**	**0.1026**	**0.0737**	**0.0429**	**0.0423**
tstat	(3.38)	(1.01)	(1.13)	(1.01)	(2.01)	(2.96)	(2.66)	(3.74)	(2.01)	(3.23)	(3.23)
ew tstat		(−0.97)	(−0.98)	(−0.97)	(−1.01)	(1.91)	(1.81)	**(−2.76)**	(−0.82)	(1.34)	(1.57)
Volatility	0.08	0.32	0.29	0.32	0.32	0.06	0.07	0.15	0.21	0.07	0.07
Sharpe ratio	0.64	0.30	0.32	0.30	0.31	0.60	0.53	0.67	0.35	0.62	0.63
Sortino ratio	1.15	0.44	0.45	0.44	0.45	0.95	0.82	0.94	0.50	1.06	1.10
Omega ratio	1.65	1.27	1.29	1.27	1.27	1.57	1.52	1.69	1.32	1.62	1.63
α		−0.0029	−0.0032	−0.0029	−0.0013	0.0045	0.0018	**0.0368**	0.0020	0.0015	0.0016
		(−0.07)	(−0.09)	(−0.07)	(−0.03)	(0.74)	(0.25)	(2.07)	(0.06)	(0.36)	(0.45)
β		**2.02**	**1.93**	**2.02**	**2.01**	**0.62**	**0.67**	**1.34**	**1.46**	**0.84**	**0.83**
		(9.19)	(9.67)	(9.18)	(9.23)	(16.91)	(19.16)	(11.07)	(8.02)	(36.54)	(38.87)
Adj. R^2		0.23	0.26	0.23	0.23	0.65	0.61	0.45	0.28	0.87	0.88

alphas of the MIN and RRT strategies relative to the corresponding EW benchmarks. Corroborating our earlier evidence, we note also that weighting schemes based on utility maximization fail to enhance performance.

Boons and Prado (2019) demonstrate that basis momentum (BM), defined as the difference between the one-year averaged returns of front and second-nearest contracts, predicts commodity futures returns. They present evidence of a positive BM risk premium captured by taking equally-weighted long positions in the contracts with the highest BM signals and equally-weighted short positions in the contracts with the lowest BM signals. Table 3, Panel C presents summary statistics for the performance of BM portfolios constructed with alternative weighting schemes. We observe that the EW scheme employed by Boons and Prado (2019) is hard to beat: the risk-adjusted performance measures of MIN, VaRT, and CVaRT are very close to those of EW. Only one scheme, RRT, with an annualized alpha of 3.68% (t-statistic of 2.07) outperforms EW. Thus, our former conclusion regarding the superiority of the risk-minimizing and risk-timing schemes is less supported in the case of BM than for the other sorting signals. However, consistent with the evidence reported before for all sorting signals, we note the poor performance of BM portfolios based on utility-maximizing weights.[11]

[11] We also study the performance of long-only portfolios formed using our 11 weighting schemes and confirm the superiority of two risk-timing schemes (RRT and BT) relative to both EW and the utility-maximizing weights. The commodity pricing literature has long recognized the superiority of being long-short rather than long-only (Miffre, 2016, for a review) and thus, to conserve space, these results are available from the authors upon request.

5.2. Turnover and transaction costs

As transaction costs can erode the profits of otherwise lucrative strategies (Do and Faff, 2012), we measure the turnover and net performance of the various momentum and term structure strategies relative to three sets of transaction costs. These include the costs of opening and closing positions as contracts enter or exit the portfolios, the costs of rolling contracts forward as they come close to maturity and; finally, the costs of rebalancing existing positions to their desired weights at the end of each holding period.[12]

Following DeMiguel et al. (2009), we measure the monthly turnover (TO) of each strategy as the average of all trades

$$TO = \frac{1}{T-1}\sum_{t=1}^{T-1}\sum_{i=1}^{N}(|x_{i,t+1} - x_{i,t^+}|), \tag{16}$$

x_{i,t^+} and $x_{i,t+1}$ are the weights assigned to the ith futures at the end of month $t+1$ before and after portfolio rebalancing, respectively; thus our definition of turnover takes into account the fact that portfolio weights naturally evolve with the performance of the assets. T is the number of observations in the holding period of the portfolios and N is the number of commodity futures present in the cross section at each month end. The turnover ranges from 0 (when no trading occurs) to 2 (when all existing positions are closed and N new positions are subsequently opened).

We also calculate net returns as in Eq. (17) assuming a conservative level of transaction costs (TC) at 8.6 bps (Marshall et al., 2012).

$$r_{p,t+1} = \sum_{i=1}^{N} x_{i,t} r_{i,t+1} - \sum_{i=1}^{N} TC \times |x_{i,t+1} - x_{i,t^+}| \tag{17}$$

We complement this investigation with a break-even analysis that calculates the level of transaction costs above which a given strategy becomes unprofitable. The higher the break-even transaction costs, the more likely the strategy is to be profitable net of reasonable transaction costs.

Table 4 reports the results of our transaction cost analysis for the momentum strategies in Panel A and for the term structure strategies in Panel B. Aligned with (Fleming et al., 2001) and Kirby and Ostdiek (2012), we note lower turnover, lower transaction costs and higher investment practicality of the risk-timing schemes relative to the utility-maximizing schemes. The estimated break-even costs range from 15 bps to 58 bps and thus are much higher than the conservative 8.6 bps estimate of Marshall et al. (2012). To state this differently, transaction costs are unlikely to wipe out the superior performance identified in Tables 1 and 2 for some of the strategies and indeed, the momentum and term structure strategies that outperformed their equal-weight counterparts in Table 1 (MIN, VT, VaRT, and CVaRT) and Table 2 (MIN, RRT, and BT) still offer superior net Sharpe ratios in Table 4. In other words, the ranking of performance is unchanged and the conclusions drawn thus far are robust to the consideration of transaction costs.

5.3. Illiquidity

As further robustness checks, we test whether the performance identified in Tables 1 and 2 is merely driven by the lack of liquidity of the constituents traded in the long-short momentum and term structure portfolios. To do so, we conduct three tests. First, we remove from the cross section available at the time of portfolio formation the 10% of contracts that have the highest Amihud measure (namely, those with the highest ratio of absolute return to $Volume as averaged over the past two months) and implement the strategies on the remaining 90%. In the same spirit, we consider two portfolios, one that considers the 80% most liquid assets and one that considers the 80% least liquid assets based on Amihud measure. If lack of liquidity drives part of the performance, the strategies that omit the most illiquid assets should exhibit a worse performance. Finally, we measure the alpha of the naive and sophisticated strategies relative to the three-factor model of Bakshi et al. (2019), augmented with a risk premium based on (Amihud et al., 1997) liquidity measure.

The results are reported in Table 5. Omitting or considering the least liquid assets has no bearing on relative performance. Besides, the measures of abnormal performance are of a similar magnitude as previously reported in Tables 1 and 2, Panels C and the exposures of the strategies to the liquidity risk premium are often equal to zero at the 5% level. We conclude therefore that the outperformance of the MIN, VT, VaRT, and CVaRT momentum strategies and that of the MIN, RRT, and BT term structure strategies relative to their naive counterparts are not merely a compensation for liquidity risk.

5.4. Is superior performance a result of data snooping?

Our conclusion thus far is that strategies that adopt weighting schemes based on risk minimization and risk timing (MIN, VT, RRT, BT, VaRT, and CVaRT) perform at least as well as strategies based on equal or utility-maximizing weights (EW, PU, NE, DA, and MV). Is this a result of data snooping? We use the Superior Predictive Ability test of Hansen (2005) to address this issue.

We treat each of the outperforming schemes (MIN, VT, RRT, BT, VaRT, and CVaRT) in turn as benchmark (b) and compare the Sharpe ratios of the five underperforming schemes (EW, PU, NE, DA, and MV) to that of the chosen benchmark. Let SR_m denote the

[12] While important, transaction costs are unlikely to completely wipe out trading profits in our context. First, compared to spot assets, futures contracts are cheap to trade. Second, the contracts traded are liquid as they are located at the front end of the futures curve. Third, unlike equities, futures contracts are as cheap to sell short as they are to buy and are not subject to short-selling ban. Finally, the strategies trade 40 contracts at most and thus they are far less trading intensive than those typically implemented in equity markets.

Table 4

Transaction costs. The table studies the impact of transaction costs on the performance of various momentum (Panel A) and term structure (Panel B) strategies. The mean turnover measures the monthly average number of trades per strategy. Net mean excess returns and net Sharpe ratios are computed after deducting from the gross returns round-trip transaction costs of 8.6 bps. Break-even transaction cost is the cost above which the strategies become unprofitable. EW, PU, NE, DA, MV, and MIN stand for equal weights, power utility, negative exponential utility, power utility with disappointment aversion, mean variance, and minimum variance, respectively. VT, RRT, BT, VaRT, and CVaRT stand for volatility timing, reward-to-risk timing, beta-timing, Value-at-Risk timing and Conditional Value-at-Risk timing, respectively. Bold fonts denote significance at the 5% level or better and t-ratios are shown in parentheses. The analysis spans the sample January 1980 to April 2016.

	EW	*Optimized weights*					*Risk-timing weights*				
		PU	NE	DA	MV	MIN	VT	RRT	BT	VaRT	CVaRT
Panel A: Momentum strategies											
Monthly turnover											
Mean	1.52	1.71	1.68	1.71	1.73	1.58	1.58	1.59	1.66	1.56	1.55
Median	1.49	1.99	1.88	1.99	1.99	1.60	1.58	1.61	1.70	1.55	1.53
Minimum	0.75	0.01	0.16	0.01	0.00	0.45	0.59	0.69	0.61	0.72	0.70
Maximum	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Net Mean return	**0.05**	0.06	0.06	0.06	0.06	**0.04**	**0.04**	**0.08**	**0.06**	**0.05**	**0.05**
	(3.80)	(1.06)	(1.22)	(1.06)	(1.35)	(4.23)	(4.04)	(3.55)	(2.39)	(4.06)	(4.08)
Net Sharpe ratio	0.55	0.16	0.18	0.16	0.17	0.70	0.67	0.57	0.34	0.65	0.65
Break-even transaction cost (bps)	33	36	36	36	37	30	32	47	40	33	33
Panel B: Term structure strategies											
Monthly turnover											
Mean	1.57	1.76	1.73	1.76	1.76	1.64	1.61	1.64	1.73	1.58	1.57
Median	1.54	2.00	1.92	2.00	2.00	1.64	1.61	1.69	1.84	1.56	1.54
Minimum	0.84	0.01	0.17	0.01	0.01	0.76	0.62	0.71	0.56	0.75	0.71
Maximum	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Net Mean return	**0.03**	0.09	0.08	0.09	0.09	**0.02**	**0.01**	**0.10**	**0.10**	**0.02**	**0.02**
	(2.65)	(0.81)	(0.90)	(0.81)	(1.79)	(3.52)	(2.06)	(3.58)	(2.94)	(2.46)	(2.46)
Net Sharpe ratio	0.38	0.26	0.27	0.26	0.26	0.43	0.15	0.72	0.53	0.32	0.32
Break-even transaction cost (bps)	26	49	48	49	50	21	15	58	57	21	21

Table 5

Liquidity risk. The table studies the impact of lack of liquidity on the performance of various momentum (Panel A) and term structure (Panel B) strategies. SR(All), SR(Liquid - n%) and SR(Illiquid - n%) are the Sharpe ratios of the strategies when implemented on the whole cross section, the n% cross section that presents the highest and lowest liquidity ratios (Amihud et al., 1997), respectively. α, $\beta(LIQ)$, Newey–West t-statistics and adjusted-R^2 are obtained from regressions of the excess returns of a given naive or sophisticated strategy on the excess returns of the three-factor benchmark of Bakshi et al. (2019) augmented with a liquidity risk premium (LIQ). Alpha has been annualized. PU, NE, DA, MV, and MIN stand for power utility, negative exponential utility, power utility with disappointment aversion, mean variance utility, and minimum variance, respectively. VT, RRT, BT, VaRT, and CVaRT stand for volatility timing, reward-to-risk timing, beta-timing, Value-at-Risk timing and Conditional Value-at-Risk timing, respectively. Bold fonts denote significance at the 5% level or better. The analysis spans the sample January 1980 to April 2016.

	Equal Weights	*Optimized weights*					*Risk-timing weights*				
		PU	NE	DA	MV	MIN	VT	RRT	BT	VaR	CVaR
Panel A: Momentum strategies											
SR (all)	0.74	0.21	0.24	0.21	0.22	0.97	0.95	0.69	0.44	0.87	0.87
SR (Liquid - 90%)	0.75	0.14	0.17	0.14	0.15	0.98	0.94	0.70	0.40	0.87	0.86
SR (Liquid - 80%)	0.68	0.24	0.24	0.24	0.25	1.02	0.89	0.70	0.35	0.81	0.81
SR (Illiquid - 80%)	0.66	0.15	0.18	0.15	0.16	0.86	0.86	0.63	0.41	0.79	0.79
α	**0.0432**	−0.0166	−0.0092	−0.0167	−0.0128	**0.0454**	**0.0540**	**0.0544**	0.0376	**0.0498**	**0.0492**
	(4.15)	(−0.30)	(−0.19)	(−0.30)	(−0.23)	(5.40)	(5.42)	(3.23)	(1.34)	(5.24)	(5.21)
$\beta(LIQ)$	−0.03	−0.18	−0.15	−0.18	−0.19	−0.05	−0.03	−0.03	0.12	−0.03	−0.03
	(−0.56)	(−0.73)	(−0.69)	(−0.73)	(−0.74)	(−1.72)	(−1.11)	(−0.36)	(0.85)	(−0.88)	(−0.81)
$Adj-R^2$	0.36	0.32	0.34	0.32	0.32	0.25	0.13	0.46	0.23	0.26	0.26
Panel B: Term structure strategies											
SR (all)	0.58	0.31	0.33	0.31	0.31	0.73	0.43	0.84	0.62	0.56	0.55
SR (Liquid)	0.52	0.32	0.32	0.32	0.32	0.62	0.36	0.81	0.51	0.49	0.48
SR (Liquid - 80%)	0.51	0.41	0.41	0.41	0.42	0.73	0.33	0.85	0.51	0.44	0.44
SR (Illiquid - 80%)	0.55	0.25	0.28	0.25	0.26	0.65	0.38	0.70	0.58	0.52	0.52
α	**0.0294**	0.0163	0.0173	0.0163	0.0180	**0.0265**	**0.0204**	**0.0687**	0.0551	**0.0249**	**0.0248**
	(3.46)	(0.32)	(0.39)	(0.32)	(0.36)	(4.09)	(2.31)	(3.46)	(1.91)	(3.01)	(3.09)
$\beta(LIQ)$	0.04	−0.36	−0.31	−0.36	−0.36	−0.00	0.03	−0.07	0.18	0.05	**0.05**
	(1.46)	(−1.72)	(−1.66)	(−1.73)	(−1.73)	(−0.01)	(1.08)	(−0.97)	(1.63)	(1.80)	(2.07)
$Adj-R^2$	0.68	0.28	0.29	0.28	0.28	0.47	0.47	0.38	0.24	0.62	0.64

Sharpe ratio of strategy $m = \{EW, PU, NE, DA, MV, b\}$ and SR_{max} be such that $SR_{max} = max(SR_{EW}, SR_{PU}, SR_{NE}, SR_{DA}, SR_{MV}, SR_b)$. Following Hansen (2005), we define three loss functions for the m strategies. One is linear, $L_m = SR_{max} - SR_m$; the other two are nonlinear, $L_m = 1/exp(\lambda SR_m)$ with curvature coefficients $\lambda = \{1, 2\}$. In all three settings, the expected loss of strategy $i = \{EW, PU, NE, DA, MV\}$ relative to the chosen benchmark b is measured as $E(d_i) = E(L_b - L_i)$. Strategy i is superior to the benchmark if $E(d_i) > 0$.

Table 6

Data snooping test. The table presents p-values for the Superior Predictive Ability test of Hansen (2005). The null hypothesis is that the benchmark is superior to the EW portfolio and to portfolios based on PU, NE, DA, and MV in terms of Sharpe ratio. MIN, VT, RRT, BT, VaRT and CVaRT stand for minimum volatility, volatility timing, reward-to-risk timing, beta-timing, Value-at-Risk timing, and Conditional Value-at-Risk timing, respectively. The analysis spans the sample January 1980 to April 2016.

Model specifications		*benchmark*					
Loss function	p	MIN	VT	RRT	BT	VaRT	CVaRT
Panel A: Momentum strategies							
$L_m = SR_{max} - SR_m$	0.2	0.50	0.49	0.50	0.50	0.51	0.51
$L_m = SR_{max} - SR_m$	0.5	0.49	0.49	0.51	0.49	0.51	0.49
$L_m = 1/exp(SR_m)$	0.2	0.49	0.50	0.50	0.50	0.50	0.49
$L_m = 1/exp(SR_m)$	0.5	0.50	0.50	0.50	0.49	0.49	0.49
$L_m = 1/exp(2SR_m)$	0.2	0.49	0.49	0.50	0.50	0.49	0.50
$L_m = 1/exp(2SR_m)$	0.5	0.49	0.50	0.50	0.50	0.49	0.49
Panel B: Term structure strategies							
$L_m = SR_{max} - SR_m$	0.2	0.42	0.44	0.43	0.42	0.42	0.41
$L_m = SR_{max} - SR_m$	0.5	0.41	0.42	0.42	0.43	0.42	0.42
$L_m = 1/exp(SR_m)$	0.2	0.41	0.42	0.41	0.42	0.42	0.41
$L_m = 1/exp(SR_m)$	0.5	0.41	0.42	0.41	0.41	0.42	0.42
$L_m = 1/exp(2SR_m)$	0.2	0.41	0.41	0.42	0.42	0.41	0.42
$L_m = 1/exp(2SR_m)$	0.5	0.41	0.42	0.41	0.42	0.42	0.43

For a given benchmark b, we use the bootstrap method of Politis and Romano (1994) to test the null hypothesis that the best of the five underperforming weighting schemes does not beat the chosen benchmark; i.e., $H_0 : E(d_i) \leq 0$ for $i = \{EW, PU, NE, DA, MV\}$. We obtain 10,000 bootstrap time-series of excess returns for the chosen benchmark and the five underperforming schemes, $\{r^*_{b,t}, r^*_{EW,t}, r^*_{PU,t}, r^*_{NE,t}, r^*_{DA,t}, r^*_{MV,t}\}$, by pooling random blocks from the original time-series of excess returns.[13] Subsequently we obtain 10,000 pseudo values for L^*_b and L^*_i for each of the $i = \{EW, PU, NE, DA, MV\}$ strategies. We set $d^* = L^*_b - min(L^*_i)$. The p-values reported in Table 6 show that the null hypothesis cannot be rejected for any loss function. The conclusion holds irrespective of the benchmark (MIN, VT, RRT, BT, VaRT, and CVaRT) considered. It is also robust to the choice of signal (momentum in Panel A and term structure in Panel B). Altogether, we conclude that the superiority of the sophisticated weighting schemes based on risk minimization and risk timing cannot be attributed to data snooping.

5.5. Choice of model parameters

Our base-case assumes an estimation period of 12 months for the portfolio weights, a 12-month ranking period for momentum, a 1-month holding period, and a tuning parameter η equal to 1. Table 7 allays concerns regarding the sensitivity of our results to these choices by considering three weighting ranking periods (R_{weight}) set to 3, 6, or 18 months,[14] two momentum ranking periods (R_{mom}) set to 3 and 6, two holding periods (H) set to 2 and 4 months, and two tuning parameters (η) set to 2 and 3. The table presents the Sharpe ratios that are obtained under these model specifications and shows that the conclusions are mostly unchanged. With only a few exceptions, the sophisticated momentum and term structure strategies that outperform their equal-weight counterparts are the same under these alternative settings as those in Tables 1 and 2. The exceptions pertain, for example, to term structure portfolios with a holding period of 4 months and to VT momentum strategies with tuning parameters η exceeding 1.

5.6. Sub-sample analysis

Finally, we appraise the performance of the strategies over various sub-periods. These are delimited as follows: pre and post financialization dated January 2006 (Stoll and Whaley, 2010; Tang and Xiong, 2012); periods of heightened versus reduced conditional volatility in commodity (or equity) markets (where the threshold is defined as the long-term average S&P-GSCI (S&P500) volatility estimated from a GARCH(1,1) model), as well as over five consecutive periods of roughly equal lengths. The conclusions presented in Table 8, Panel A for the momentum strategies are remarkably consistent with those reported in Table 1: irrespective of the sample considered, the Sharpe ratios of the MIN, VT, VaRT, and CVaRT exceed those of EW. The inferences presented in Table 8, Panel B for the term structure strategies are slightly less robust. As in Table 2, MIN and RRT nearly always generate higher Sharpe ratios than EW but that conclusion applies less often to BT. Interestingly, the strategies perform better in period of low volatility in commodity and equity markets, suggesting that the timing strategies of Barroso and Santa-Clara (2015) could enhance performance further.

[13] The length of each sample block follows a geometric distribution with expected value $1/p$ with $p = \{0.2, 0.5\}$.

[14] The use of a longer estimation period shall mitigate legitimate concerns regarding estimation errors.

Table 7

Model parameters. The table studies the sensitivity of the Sharpe ratios of various momentum (Panel A) and term structure (Panel B) strategies to the choice of weighting ranking periods (R_{weight}), momentum ranking periods (R_{mom}), holding period (H), and tuning parameter (η). PU, NE, DA, MV, and MIN stand for power utility, negative exponential utility, power utility with disappointment aversion, mean variance, and minimum variance, respectively. VT, RRT, BT, VaRT, and CVaRT stand for volatility timing, reward-to-risk timing, beta-timing, Value-at-Risk timing and Conditional Value-at-Risk timing, respectively.

	Equal	*Optimized weights*					*Risk-timing weights*				
	Weights	PU	NE	DA	MV	MIN	VT	RRT	BT	VaR	CVaR
Panel A: Momentum-based strategies											
$R_{weight} = 3$	0.74	0.17	0.17	0.17	0.18	0.91	0.79	0.71	0.54	0.87	0.69
$R_{weight} = 6$	0.74	0.12	0.15	0.12	0.13	1.03	0.86	0.68	0.52	0.84	0.85
$R_{weight} = 18$	0.76	−0.11	−0.06	−0.11	−0.10	0.96	0.92	0.36	0.40	0.85	0.87
$R_{mom} = 3$	0.86	0.36	0.39	0.36	0.37	0.81	0.94	0.79	0.47	0.94	0.91
$R_{mom} = 6$	0.65	0.20	0.22	0.20	0.21	0.84	0.77	0.67	0.35	0.77	0.76
$H = 2$	0.69	0.11	0.14	0.11	0.32	0.94	0.78	0.56	0.40	0.76	0.77
$H = 4$	0.35	0.10	0.10	0.10	0.11	0.50	0.41	0.34	0.24	0.39	0.40
$\eta = 2$							0.73	0.56	0.44	0.89	0.92
$\eta = 4$							0.53	0.48	0.45	0.75	0.85
Panel B: Term structure-based strategies											
$R_{weight} = 3$	0.36	0.05	0.05	0.05	0.05	0.40	0.28	0.65	0.51	0.41	0.43
$R_{weight} = 6$	0.40	−0.05	−0.03	−0.05	−0.05	0.46	0.29	0.63	0.54	0.40	0.39
$R_{weight} = 18$	0.47	0.08	0.11	0.08	0.08	0.60	0.34	0.38	0.41	0.40	0.41
$H = 2$	0.50	0.24	0.26	0.24	0.24	0.69	0.40	0.68	0.54	0.51	0.50
$H = 4$	0.54	0.18	0.19	0.18	0.19	0.58	0.33	0.37	0.42	0.45	0.46
$\eta = 2$							0.19	0.66	0.64	0.47	0.49
$\eta = 4$							0.08	0.55	0.64	0.32	0.39

Table 8

Sub-sample analysis. The table studies the sensitivity of the Sharpe ratios of various momentum (Panel A) and term structure (Panel B) strategies to different subsamples. PU, NE, DA, MV, and MIN stand for power utility, exponential utility, power utility with disappointment aversion, mean variance, and minimum variance, respectively. VT, RRT, BT, VaRT, and CVaRT stand for volatility timing, reward-to-risk timing, beta-timing, Value-at-Risk timing, and Conditional Value-at-Risk timing, respectively. Periods of heightened versus reduced volatility in commodity (or equity) markets are defined relative to the long-term average S&P-GSCI (S&P500) conditional volatility estimated from a GARCH(1,1) model.

	Equal	*Optimized weights*					*Risk-timing weights*				
	Weights	PU	NE	DA	MV	MIN	VT	RRT	BT	VaR	CVaR
Panel A: Momentum strategies											
Pre financialization	0.98	0.26	0.28	0.26	0.27	1.18	1.15	0.81	0.71	1.12	1.13
Post financialization	0.29	0.08	0.10	0.08	0.08	0.63	0.63	0.40	−0.16	0.48	0.45
Low S&P-GSCI volatility	1.01	0.60	0.62	0.60	0.61	1.27	1.15	0.99	0.76	1.15	1.12
High S&P-GSCI volatility	0.52	−0.19	−0.16	−0.19	−0.18	0.79	0.79	0.46	0.22	0.66	0.67
Low S&P-500 volatility	0.85	0.31	0.35	0.31	0.32	1.12	1.07	0.90	0.58	1.02	0.97
High S&P-500 volatility	0.64	0.07	0.09	0.07	0.08	0.81	0.81	0.49	0.29	0.73	0.77
Jan 80–Mar 87	0.71	0.16	0.16	0.16	0.18	1.15	0.92	0.46	0.15	0.84	0.87
Apr 87–Jun 94	0.95	0.74	0.77	0.74	0.74	1.06	1.14	0.95	0.72	1.08	1.10
Jul 94–Sep 01	1.64	0.29	0.33	0.29	0.29	2.27	1.85	1.24	1.23	1.90	1.87
Oct 01–Dec 08	0.68	−0.06	0.01	−0.06	−0.04	0.86	0.99	0.84	0.32	0.85	0.85
Jan 09–Apr 16	0.02	−0.12	−0.12	−0.12	−0.12	0.37	0.39	0.09	−0.22	0.22	0.19
Panel B: Term structure strategies											
Pre financialization	0.79	0.36	0.38	0.36	0.36	0.83	0.64	1.00	0.98	0.81	0.80
Post financialization	0.12	0.17	0.19	0.17	0.17	0.55	0.01	0.49	−0.19	0.08	0.08
Low S&P-GSCI volatility	0.76	0.62	0.64	0.62	0.62	0.87	0.65	1.00	0.92	0.77	0.76
High S&P-GSCI volatility	0.43	−0.04	−0.01	−0.04	−0.04	0.63	0.26	0.69	0.44	0.39	0.39
Low S&P-500 volatility	0.61	0.40	0.43	0.40	0.40	0.76	0.45	0.95	0.83	0.56	0.51
High S&P-500 volatility	0.57	0.16	0.19	0.16	0.17	0.70	0.42	0.70	0.38	0.57	0.61
Jan 80–Mar 87	0.79	0.74	0.73	0.74	0.75	0.71	0.67	0.94	0.72	0.78	0.85
Apr 87–Jun 94	0.92	0.56	0.61	0.56	0.57	0.98	1.34	1.14	1.15	1.16	1.10
Jul 94–Sep 01	1.27	0.29	0.31	0.29	0.29	1.50	0.81	1.22	1.11	1.34	1.22
Oct 01–Dec 08	0.26	0.00	0.02	0.00	0.02	0.39	0.07	1.01	0.45	0.19	0.18
Jan 09–Apr 16	−0.05	−0.13	−0.11	−0.13	−0.13	0.49	−0.11	0.02	−0.21	−0.10	−0.08

6. Conclusions

This study introduces a new long-short portfolio construction technique that simultaneously captures the fundamentals of backwardation and contango present in commodity futures markets and allows for weighting schemes that depart from the traditional naive weights. These sophisticated weighting schemes emanate either from the optimization literature as first developed by Markowitz (1952) or from the risk-timing literature as epitomized by Kirby and Ostdiek (2012) (VT, RRT, and BT) and as subsequently extended in the present paper (VaRT and CVaRT). By comparing the out-of-sample performance of the sophisticated long-short momentum and term structure strategies to that of their naive equal-weight counterparts, we test whether sophistication of the weighting scheme has bearing on performance.

The empirical results show that the equal-weight scheme that is traditionally employed in the construction of long-short portfolios can be beaten. Momentum and term structure portfolios that use weights that either minimize or time risk produce greater risk-adjusted performance than the naive equal-weight portfolios. At the other end of the spectrum, weighting schemes based on utility maximization lead to poorly diversified portfolios with weaker risk-adjusted performance. These conclusions persist after accounting for the risks of the strategies as captured by the now traditional commodity pricing model of Bakshi et al. (2019), and are broadly unaltered by the consideration of alternative signals based on hedging pressure, speculative pressure, or basis-momentum, transaction costs, illiquidity, data snooping, various model specifications, or different sub-periods. We attribute the better out-of-sample performance of the risk-timing schemes to lower estimation errors and reasonable turnover and the poor performance of the utility-maximizing schemes to estimation risk, extreme long-short weights and high turnover. By doing so, our paper extends to a long-short setting the conclusions that Kirby and Ostdiek (2012) drew in a long-only context.

CRediT authorship contribution statement

Hossein Rad: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing. **Rand Kwong Yew Low:** Conceptualization, Methodology, Software, Writing - review & editing. **Joëlle Miffre:** Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Robert Faff:** Writing - review & editing.

Appendix. Summary statistics for the risk factors

The table presents summary statistics for the excess returns of long-only and long-short portfolios. AVG is a long-only equally-weighted monthly-rebalanced portfolio of all commodity futures, S&P-GSCI stands for Standard & Poor's Goldman Sachs Commodity Index, CARRY, MOM and LIQ are equally-weighted monthly-rebalanced long-short portfolios based on roll yields, past performance and liquidity ratios, respectively. Mean, volatility and downside volatility have been annualized, significance t-ratios for the annualized mean excess returns are shown in parentheses. Bold fonts denote significance at the 5% level or better. The analysis spans the sample January 1980 to April 2016.

	Long-only portfolios		*Long-short portfolios*		
	AVG	S&P-GSCI	CARRY	MOM	LIQ
Mean	0.0169	0.0341	**0.0652**	**0.0520**	**0.0569**
tstat	(0.68)	(0.83)	(3.00)	(2.99)	(2.87)
Volatility	0.12	0.20	0.12	0.13	0.11
Downside volatility	0.09	0.14	0.08	0.08	0.06
Skewness	**−0.61**	**−0.31**	−0.20	−0.05	**0.36**
Ex. kurtosis	**2.88**	**2.24**	**2.82**	**0.75**	**0.93**
JB p.value	0.00	0.00	0.00	0.00	0.00
VaR (95%)	−0.05	−0.10	−0.05	−0.06	−0.04
CVaR (95%)	−0.08	−0.13	−0.07	−0.08	−0.06
Max drawdown	0.31	0.52	0.27	0.27	0.22
Sharpe ratio	0.14	0.17	0.57	0.40	0.53
Sortino ratio	0.19	0.24	0.83	0.64	0.96
Omega ratio	1.11	1.14	1.56	1.35	1.51

References

Amihud, Y., Mendelson, H., Lauterbach, B., 1997. Market microstructure and securities values: evidence from the Tel Aviv Stock Exchange. J. Financ. Econ. 45 (3), 365–390.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. Math. Finance 9 (3), 203–228.

Asness, C.S., Moskowitz, T.J., Pedersen, L.H., 2013. Value and momentum everywhere. J. Finance 68 (3), 929–985.

Bakshi, G., Gao, X., Rossi, A.G., 2019. Understanding the sources of risk underlying the cross section of commodity returns. Manage. Sci. 65 (2), 619–641.

Barroso, P., Santa-Clara, P., 2015. Momentum has its moments. J. Financ. Econ. 116 (1), 111–120.

Basu, D., Miffre, J., 2013. Capturing the risk premium of commodity futures: the role of hedging pressure. J. Bank. Financ. 37 (7), 2652–2664.

Bessembinder, H., 1992. Systematic risk, hedging pressure, and risk premiums in futures markets. Rev. Financ. Stud. 5 (4), 637–667.

Boons, M., Prado, M.P., 2019. Basis-Momentum. J. Finance 74 (1), 239–279.

Cootner, P.H., 1960. Returns to speculators: Telser versus Keynes. J. Political Economy 68 (4), 396–404.

Daskalaki, C., Skiadopoulos, G., 2011. Should investors include commodities in their portfolios after all? New evidence. J. Bank. Financ. 35 (10), 2606–2626.

De Roon, F.A., Nijman, T.E., Veld, C., 2000. Hedging pressure effects in futures markets. J. Finance 55 (3), 1437–1456.

DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy?. Rev. Financ. Stud. 22 (5), 1915–1953.

Do, B., Faff, R., 2012. Are pairs trading profits robust to trading costs?. J. Financ. Res. 35 (2), 261–287.

Driessen, J., Maenhout, P., 2007. An empirical portfolio perspective on option pricing anomalies. Rev. Finance 11 (4), 561–603.

Erb, C.B., Harvey, C.R., 2006. The strategic and tactical value of commodity futures. Financ. Anal. J. 62 (2), 69–97.

Fama, E.F., French, K.R., 1987. Commodity futures prices: some evidence on forecast power, premiums, and the Theory of Storage. J. Bus. 60 (1), 55–73.
Fernandez-Perez, A., Frijns, B., Fuertes, A.-M., Miffre, J., 2018. The skewness of commodity futures returns. J. Bank. Financ. 86, 143–158.
Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing. J. Finance 56 (1), 329–352.
Gao, X., Nardari, F., 2018. Do commodities add economic value in asset allocation? New evidence from time-varying moments. J. Financ. Quant. Anal. 53 (1), 365–393.
Gorton, G.B., Hayashi, F., Rouwenhorst, K.G., 2013. The fundamentals of commodity futures returns. Rev. Finance 17 (1), 35–105.
Gorton, G., Rouwenhorst, K.G., 2006. Facts and fantasies about commodity futures. Financ. Anal. J. 62 (2), 47–68.
Gul, F., 1991. A theory of disappointment aversion. Econometrica 59 (3), 667–686.
Hansen, P.R., 2005. A test for superior predictive ability. J. Bus. Econom. Statist. 23 (4), 365–380.
Hirshleifer, D., 1990. Hedging pressure and futures price movements in a general equilibrium model. Econometrica 58 (2), 411–428.
Hong, H., Yogo, M., 2012. What does futures market interest tell us about the macroeconomy and asset prices?. J. Financ. Econ. 105 (3), 473–490.
Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: why Imposing the wrong constraints helps. J. Finance 58 (4), 1651–1683.
Kaldor, N., 1939. Speculation and economic stability. Rev. Econom. Stud. 7 (1), 1–27.
Kandel, S., Stambaugh, R.F., 1996. On the predictability of stock returns: an asset-allocation perspective. J. Finance 51 (2), 385–424.
Kang, W., Rouwenhorst, K.G., Tang, K., 2020. A tale of two premiums: the role of hedgers and speculators in commodity futures markets. J. Finance 75 (1), 377–417.
Kirby, C., Ostdiek, B., 2012. It's all in the timing: simple active portfolio strategies that outperform naive diversification. J. Financ. Quant. Anal. 47 (02), 437–467.
Kolm, P.N., Tutuncu, R., Fabozzi, F.J., 2014. 60 years of portfolio optimization: practical challenges and current trends. European J. Oper. Res. 234 (2), 356–371.
Kostakis, A., Panigirtzoglou, N., Skiadopoulos, G., 2011. Market timing with option-implied distributions: a forward-looking approach. Manage. Sci. 57 (7), 1231–1249.
Low, R.K.Y., Alcock, J., Faff, R., Brailsford, T., 2013. Canonical vine copulas in the context of modern portfolio management: are they worth it?. J. Bank. Financ. 37 (8), 3085–3099.
Low, R.K.Y., Faff, R., Aas, K., 2016. Enhancing mean–variance portfolio selection by modeling distributional asymmetries. J. Econ. Bus. 85, 49–72.
Markowitz, H., 1952. Portfolio selection. J. Finance 7 (1), 77–91.
Marshall, B.R., Nguyen, N.H., Visaltanachoti, N., 2012. Commodity liquidity measurement and transaction costs. Rev. Financ. Stud. 25 (2), 599–638.
Merton, R.C., 1980. On estimating the expected return on the market: an exploratory investigation. J. Financ. Econ. 8 (4), 323–361.
Michaud, R.O., 2008. Efficient asset management: a practical guide to stock portfolio optimization and asset allocation. In: OUP Catalogue, Oxford University Press.
Miffre, J., 2016. Long-short commodity investing: A review of the literature. J. Commod. Mark. 1 (1), 3–13.
Miffre, J., Rallis, G., 2007. Momentum strategies in commodity futures markets. J. Bank. Financ. 31 (6), 1863–1886.
Moskowitz, T.J., Ooi, Y.H., Pedersen, L.H., 2012. Time series momentum. J. Financ. Econ. 104 (2), 228–250.
Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. J. Amer. Stat. Assoc. 89 (428), 1303–1313.
Rockafellar, R.T., Uryasev, S., 2002. Conditional Value-at-Risk for general loss distributions. J. Bank. Financ. 26 (7), 1443–1471.
Stoll, H.R., Whaley, R.E., 2010. Commodity index investing and commodity futures prices. J. Appl. Finance 20 (1), 7.
Symeonidis, L., Prokopczuk, M., Brooks, C., Lazar, E., 2012. Futures basis, inventory and commodity price volatility: An empirical analysis. Econ. Model. 29 (6), 2651–2663.
Szakmary, A.C., Shen, Q., Sharma, S.C., 2010. Trend-following trading strategies in commodity futures: a re-examination. J. Bank. Financ. 34 (2), 409–426.
Szymanowska, M., De Roon, F., Nijman, T., Van Den Goorbergh, R., 2014. An anatomy of commodity futures risk premia. J. Finance 69 (1), 453–482.
Tang, K., Xiong, W., 2012. Index investment and the financialization of commodities. Financ. Anal. J. 68 (6), 54–74.
Tu, J., Zhou, G., 2011. Markowitz meets Talmud: a combination of sophisticated and naive diversification strategies. J. Financ. Econ. 99 (1), 204–215.