Pairs Trading of Chinese and International Commodities

Adrian Fernandez-Perez – Auckland University of Technology*

Bart Frijns – Auckland University of Technology

Ivan Indriawan – Auckland University of Technology

Yiuman Tse – University of Missouri – St. louis

This version Date: November 2018

Abstract

We investigate the profitability of a pairs trading strategy using Chinese and International commodity futures contracts covering the period January 2004 to February 2018. We use a time-series approach where the commodity pairs have been pre-determined given similar underlying. Applying this strategy to a portfolio of commodities yields 2.08% annual excess returns and a Sharpe ratio of 0.79. For a portfolio of metal futures, this strategy yields 5.32% excess return and a Sharpe ratio of 1.47, whereas for gold-only futures, this strategy yields 7.39% excess returns and 1.95 Sharpe ratio. This performance is superior to traditional strategies based on term structure, momentum and value portfolios. Arbitrage opportunities in these commodity pairs remain even after accounting for transaction costs and they are robust to data snooping bias.

JEL Classification: G11, G15, G19

Keywords: Pairs Trading, Chinese Commodity Futures, Portfolio Gains

*Corresponding Author. Department of Finance, Auckland University of Technology, Auckland, New Zealand. Tel.: (+64)9 921-9999 ext.6129. *E-mail address*: adrian.fernandez@aut.ac.nz.

1. Introduction

The seminal paper by Gatev, Goetzmann, and Rouwenhorst (2006, henceforth GGR) is one of the first comprehensive examinations of the profitability of a simple pairs trading strategy in the US stocks. Their strategy identifies pairs of stocks whose prices have moved together historically over a specific formation period. If prices diverge, one shorts the winner and buy the loser, assuming that there is an equilibrium relation between the two securities and mean reversion in price spread will take place. This concept of univariate pairs trading can then be extended into a multivariate framework, i.e. groups of stocks. This strategy is often referred as generalized pairs trading, or statistical arbitrage.

This paper is related to the profitability of pairs trading strategies in the Chinese futures markets. China is currently the world's second largest economy and is the world's largest consumer of a wide range of commodities (Fung et al., 2013; Liu et al. 2018). Given the trading activity in this market, even the slightest profit margin is considerably large compared to the profit attainable by trading in many other markets. However, finding a strategy that works in the Chinese market is not a simple task because market imperfections in China can invalidate any profitable trading strategy. The Chinese commodity markets, particularly the agricultural products, are highly regulated by the government, and foreign investors have limited access to these commodities. As discussed by Liu et al. (2018) and others, the lack of institutional investors has caused China's futures markets to be dominated by retail investors and domestic financial speculators. As a result, Chinese commodities can be segmented from the world markets. The potential profitable opportunities of pairs trading between the Chinese and world prices remain an important issue that can only be examined through a detailed empirical analysis.

In this study, we examine the profitability of pairs trading strategy using Chinese and International commodity futures contracts. Our sample includes ten commodity pairs - aluminum, copper, gold, zinc, natural rubber, wheat, cotton, soybeans, soymeal and corn – with data covering a period of more than 10 years, from January 2004 to February 2018.[1] We first examine whether pairs trading in these futures yields significant positive returns. We then compare the profitability of this strategy with other traditional trading strategies (such as term-structure, momentum, and value-based portfolios) that are known to be profitable in commodity futures.

Applying the strategy to a single commodity, we obtain excess returns ranging from 7.39% p.a. with a Sharpe ratio of 1.95 in the case of gold futures to -1.80% p.a. in the case of wheat futures. We also find that metal futures as a group, offer better performance than agricultural futures with a 5.32% annual excess returns and a Sharpe ratio of 1.47. These results are consistent with the facts that compared with agricultural products in China, metals (particularly gold) in China are almost identical to the corresponding world markets and are less regulated. When we consider a portfolio of commodities, our strategy yields an excess return of 2.08% p.a. with a Sharpe ratio of 0.79. This performance is superior to traditional strategies based on term structure, momentum and value portfolios, and persists even after accounting for transaction costs, and controlling for a data snooping bias.

This paper broadens the literature of pairs trading strategies in Chinese commodities markets through the inclusion of International commodities futures. Existing literature focuses on pairs trading strategy among commodities from the same market. For instance, Chen et al. (2017) and Yang et al. (2017) examine the profitability of different pairs selection using Chinese commodity futures. They find that the profitability of these strategies depends on the identification

[1]Zinc and gold futures start in April 2007 and January 2008, respectively.

of suitable pairs. Our study differs from theirs as we pair Chinese with International commodities sharing intrinsically similar product, e.g., Chinese and UK aluminum futures.

Another contribution of this paper is in showing that statistical arbitrage strategy exists using commodities on a global scale in recent years. Do and Faff (2010, 2012) and others have reported declining profits of pairs trading in the US in more recent times (specifically when considering the time period following the study of GGR). The main reason for this decline in profitability is the lack of convergence in the spread between the two prices of a pair, indicating no equilibrium relation between the prices that were statistically paired during the formation period. In contrast to previous studies (Galenko et al, 2012, Rad et al., 2016) which relies on statistical analyses to determine a security pair (e.g. based on correlation, price minimization, and more advanced approaches using cointegration and copula), our identification of trading pairs is simple and offers economic intuition; the commodities share the same underlying. Our results show that using a simple identification method, it is possible to obtain profitable global arbitrage opportunities. Hence, our results complement Do and Faff (2010) who find that matching pairs within the same industry lowers the divergence risk and enhances trading profits.

The remainder of this paper proceeds as follows. Section 2 explains the pairs trading methodology. Section 3 presents the data sources and summary statistics. Section 4 discusses the empirical findings. Section 5 presents the robustness tests. Section 6 concludes.

2. Methodology

Several strategies based on the concept of pairs trading have been proposed in the literature. One of the most referenced pairs trading strategies, due to its simplicity, is that of GGR. They use a distance metric to identify comoving securities across different markets, asset classes or time frames. Since the focus is distance between prices, the pair may come from different assets or markets. Another commonly used strategy is the time-series approach of Elliott, Van Der Hoek and Malcolm (2005, henceforth EVM). This approach assumes that a set of co-moving securities is intrinsically linked, so the securities pair has been pre-determined during the formation period, e.g. cross-listed stocks.[2]

In this paper, we follow the second approach for several reasons. First, theoretically, this approach is more suitable in our setting because the futures in our sample have similar underlyings. They are simply traded in different futures markets so investors may consider these futures contracts as direct substitutes. As such, the securities are already in return parity and fulfil the pre-requisites of the EVM strategy better than GGR. Second, the model is completely tractable, with its parameters easily estimated by the Kalman filter in a state space setting. The estimator is based on the maximum likelihood and optimal in terms of minimum mean squared error.

In this section, we start by discussing the formation period in which we normalize price series to make them comparable. Next, we proceed with the trading period in which we specify the cut-off point when the strategy is triggered. Finally, we explain how we measure the performance of this strategy.

2.1. Formation period

During the formation period (R), we construct a normalized price series to ensure they are comparable across commodities. We normalize the price series by constructing a cumulative (log) returns series for each commodity i at time t, $p_{i,t}$. We then calculate the difference in cumulative

---

[2] See Krauss (2017) for a detailed review about the pairs trading strategies.

returns between the Chinese and International commodity pair i, i.e. $y_{i,t} = p_{i,t}^{Chn} - p_{i,t}^{Int}$. Subsequently, we calculate the mean μ, standard deviation σ, and the strength of the mean-reversion ρ, of $y_{i,t}$ using a state-space model based on Kalman Filter. Specifically, the state-space model of EVM can be expressed as follows:

The state process: $$x_{i,t+1} = a + b \cdot x_{i,t} + \sigma \cdot \varepsilon_{i,t+1} \quad (1)$$

The observation process: $$y_{i,t} = x_{i,t} + d \cdot w_{i,t} \quad (2)$$

where $x_{i,t}$ is the *latent* mean-reversion process of $y_{i,t}$, a, $b = (1 - \rho)$, σ, and d, are the parameters to estimate. The mean of the state-space model is $\mu = a/\rho$, the standard deviation is σ and the mean reversion parameter ρ. There is mean-reversion in the price differential of commodity pair i if $0 < b < 1$. ε and w are two i.i.d. independent Gaussian $N(0,1)$ processes. We estimate the parameters $\theta = \{a, b, \sigma, d\}$ using the expectation-maximization (EM) algorithm via the Kalman Filter as in EVM.

2.2. Trading period

Using the information from the formation period, we can proceed by taking long and short positions over the trading period (*H*). A trading signal is triggered, and positions are opened when the commodity pair, $y_{i,t}$, diverges from the state-space mean by c standard deviations, i.e. $\mu \pm c \cdot \left(\sigma/\sqrt{2\rho}\right)$. Specifically, when $y_{i,t} \leq \mu - c \cdot \left(\sigma/\sqrt{2\rho}\right)$, we take a long position in the Chinese futures and a short position in the International futures contracts and hold them for one period. Alternatively, when $y_{i,t} \geq \mu + c \cdot \left(\sigma/\sqrt{2\rho}\right)$, we take a short position in the Chinese futures and a long position in the International futures contracts and hold them for one period. If the signal is not triggered, i.e., $\mu - c \cdot (\sigma/\sqrt{2\rho}) < y_{i,t} < \mu + c \cdot (\sigma/\sqrt{2\rho})$, we do not take any positions. This process is repeated for each observation until the end of the trading period, *H*. The above step is repeated for the new formation and trading period.

EVM do not provide any indication about how the parameter *c* should be determined. Instead of choosing an arbitrary parameter, we optimize it in the formation period. Specifically, using the optimal parameters, $\hat{\theta}$, estimated in the formation period, we employ a grid between 0 to 10 with steps of 0.1 for *c* and maximize the in-sample Sharpe ratio of the commodity pair i, i.e. maximizing the Sharpe ratio of the commodity pair i returns for different values of *c*, obtained as

$$r_{i,t^*+1} = \left(\frac{1}{2}\right) \sum_{i=1}^{2} s_{i,t^*} \cdot r_{i,t^*+1} \tag{3}$$

with $t^* = 1, \ldots, t-1$ (formation period), $s_{i,t^*} = \{-1, 0, +1\}$ defines the short, null and long positions, respectively, according to EVM pairs trading strategy, and r_{i,t^*} the return of the commodity i. We employ the optimal $c_{i,t}^*$ for the commodity pair i over the trading period, defined above.[3] In this paper, we use expanding windows in the formation period (*R*) with an initial window of five years (or 260 weeks)[4] and a trading period (*H*) of one month (or 4 weeks).

2.3. Portfolio excess returns

---

[3]The optimal $c_{i,t}^*$ differs per commodity pair in each formation period with an average of 1 and a standard deviation of 0.79 across commodities and formation periods. We also fixed $c = 2$ and compare it with the optimal $c_{i,t}^*$. The number of open positions is considerably reduced when we employ a pre-determined c. However, the qualitative results are similar to those with the optimal $c_{i,t}^*$.

[4]The 5-year initial window is arbitrary but common in the literature. Likewise, the commodity futures literature usually employ holding periods of one month (see e.g., Fernandez-Perez et al., 2018; or Fan and Zhang, 2018). We have tried with another windows and the results are qualitative similar. These results are available from the authors.

Because positions may open and close several times during the one-month trading period, the calculation of the excess return on a portfolio of commodity pairs is a not trivial. Pairs that open and converge during the trading period will have positive cash flows. In addition, because pairs can reopen after initial convergence, they can have multiple positive cash flows during the trading period. We may also have no cash flow if prices never diverge by more than *c* standard deviations during the trading period. Pairs that open but do not converge will only have a cash flow at the end of the trading period when all positions are closed out. Therefore, the payoffs to pairs trading strategies consider reinvesting and compounding the cash flows received during the trading period.

Consider a number of *N* commodities in the pairs trading strategy. At the end of each formation period, we construct an equal-weighted portfolio using these *N* commodities. The portfolio excess return over the next month is computed as the reinvested payoffs over that period. The portfolio will again be rebalanced to equal weights at the end of the new formation period month. As such, the excess return of the portfolio during the trading period *H*, can be calculated as follows

$$r_{p,t+1} = z_t \cdot \left(\sum_{i=1}^{N} w_{i,t} \cdot s_{i,t} \cdot r_{i,t+1}\right) \tag{4}$$

where $r_{i,t}$ is the weekly (log) return of commodity i in week *t*, $s_{i,t} = \{-1, 0, +1\}$ is the position according to EVM pairs trading strategy, and z_t is a scalar that ensures the sum of the weights is 1. $w_{i,t} \equiv w_{i,t-1} \times e^{s_{i,t-1} \cdot r_{i,t}}$ is the portfolio weight that captures the mechanical evolution of the weights due to price movements during the month. The weekly returns are then compounded into monthly returns. Equation (4) therefore has a simple interpretation of a buy and hold strategy applied over the trading period.

We follow the conservative approach of GGR and measure excess returns based on committed capital, i.e. we invest an equal amount of capital in all *N* commodity futures even if the pairs trading strategy recommends not to trade on commodity pair *i*. If a pair does not trade for the whole of the trading period, we cannot allocate that capital into other pairs. Instead, we simply leave that position closed. This conservative approach can be considered as a lower bound of the potential benefits of this pairs trading strategy aiming to exploit arbitrage opportunities in global commodity futures markets.

3. Data and summary statistics

We focus on ten metal and agricultural commodities futures pairs: aluminum, copper, gold, zinc, natural rubber, wheat, cotton, soybeans, soymeal, and corn. Our sample period starts in January 2004 (April 2007 and January 2008 for zinc and gold, respectively) and ends in February 2018. These futures are traded in the Chinese futures markets as well as internationally in the US, UK, and Japan. We include only commodities that have data from our data vendor, Commodity Systems Inc. (CSI), for at least 10 years.

We focus on the most active contracts (in terms of trading volume and open interest) and roll-over to the next-active contracts within the delivery month.[5] Since time zones and trading hours differ between the Chinese and International markets, we employ weekly returns to mitigate the impact of non-synchronicity. Weekly returns are calculated as the log differences in prices (all in USD) between consecutive Wednesdays.[6] These returns can therefore be considered excess

---

[5] We use settlement prices of the most liquid contracts, which in the case of the International futures are usually the front-end contracts, but for the Chinese futures they are either the second- or third-nearby contracts.

[6] We work with Wednesday prices as the number of holidays that fall on a Wednesday is lower than any other day of the week. In the case when a public holiday falls on a Wednesday, we use the next day's price.

returns based on fully collateralized positions (see e.g., Koijen et al. 2018). In total, we have 715 weekly observations in our sample.

Table 1 presents summary statistics for the futures contracts in our sample. It reports the security symbols, exchanges and returns statistics both for the Chinese and International futures contracts. The excess returns in the Chinese futures markets range from -5.61% (natural rubber) to 12.83% (copper) p.a. However, these returns are not significantly different from zero. Similarly, excess returns for the International commodity futures range from -8.11% (natural rubber) to 10.54% (soymeal), but are not significantly different from zero. This is a common observation for commodity futures (particularly, Erb and Harvey (2006) and Gorton and Rouwenhorst (2006) document that the average annualized excess return of the individual commodity futures is approximately zero). However, portfolio strategies in commodities futures have been rewarded with positive mean returns.

INSERT TABLE 1 HERE

The pairwise correlations between the excess returns of each pre-determined commodity pair are plotted on Figure 1. As expected, the correlations between these commodity pairs are high (0.60 on average), and significant at the 1% level for all pairs. The lowest correlation is for the wheat pair, followed by corn. China is the world's top wheat producer and the wheat price is more regulated than other commodities (such as copper and soybeans as reported by Fung et al., 2003). In particular, the Chinese government buys wheat from farmers at a fixed price assigned by the government when the market price drops below that level (Patton, 2017). Although the government no longer uses similar programs for other agricultural products, the government still provides significant subsidies for farmers growing corn. These government policies explain why wheat and corn prices may not converge to world prices. Another reason for the low correlation of the wheat pair may be the different kinds of wheats traded in China and the US.[7] These two reasons can also explain why the metal pairs (particularly, gold) have the higher correlations. Metal products are not subsidized by the Chinese government and are almost identical products in other countries. As shown in the next section, these simple correlation results shed lights on the performance of the pairs trading strategies for individual pairs with gold being the best and wheat the worst.

INSERT FIGURE 1 HERE

As an example of the pairs trading strategy, Figure 2 plots the weekly normalized prices for aluminum. The bold line represents the futures contract in China whereas the grey line represents the aluminum futures contract in the UK. In general, the normalized prices track each other closely, but there are periods in which the difference between the two prices is greater than the pairs trading threshold. In such a case, we indicate the open and close positions of the EVM pairs trading strategy using the variable on the right-hand axis. There are several occurrences where the pairs trading strategy is executed, such as during 2009-2010 as well as 2012 and 2015.

INSERT FIGURE 2 HERE

4. Results

[7] We use hard red winter (HRW) wheat futures in the US and strong gluten wheat futures in China. They are close substitutes, but not the same products. We also use soft red winter (SRW) futures in the US and the results do not change qualitatively.

The state space model is estimated for each formation period and commodity pair. We report the results for the full sample in Table 2. The estimated parameters are for each commodity pair. All commodity pairs show mean-reversion as b is always between zero and one. However, deviations from the long-run mean are very persistent since b is close to one. The commodity pairs with a stronger mean-reversion for the full sample are aluminum, gold, zinc, cotton and soybeans, all of them with $b < 0.99$.

INSERT TABLE 2 HERE

Table 3 reports the results for the pairs trading strategies together with the results for metal (aluminum, copper, gold, and zinc) and agricultural commodity (natural rubber, wheat, cotton, soybeans, soymeal, and corn) futures, and for each commodity pair over the sample from February 2009 to February 2018.[8] The EVM pairs trading strategy offers a positive mean return of 2.08% p.a. significant at the 5% level, and an annual Sharpe ratio of 0.79. The certainty equivalent return (CER) which represents the risk-free return that an investor is willing to accept instead of engaging in a particular risky portfolio strategy based on power utility preferences with $\gamma = 5$, is 1.90% p.a.[9] This strategy has a low crash risk profile (e.g. maximum drawdown of -0.0519) and positive monthly returns in 59% of all months. By sector, we find that metal futures offer better performance than agricultural futures with a 5.32% annual excess returns and a Sharpe ratio of

---

[8] Note we use the first 5 years of our sample to obtain the initial signals in the EVM pairs trading strategy.

[9] The power utility certainty-equivalent-return is given by $CER = \left(\frac{12}{T}\right)\sum_{t=0}^{T-1}\frac{(1+r_{P,t+1})^{1-\gamma}-1}{1-\gamma}$ with $r_{P,t+1}$ the pairs trading portfolio excess return on month t+1 and T the number of out-of-sample months. $CER > 0$ implies that investing in the risky portfolio is more attractive than investing in a risk-free asset.

1.47. These results can be attributed to the fact that metals (gold in particular) in China are less regulated than their agricultural counterparts, leading to higher potential for profit.

With regard to the commodity pairs, all of them obtain positive mean returns except for the natural rubber, wheat and cotton pairs which have negative but insignificant mean returns. The commodity pairs with the highest Sharpe ratios are, in descending order, gold, zinc, copper and aluminum. Many individual commodity pairs are not very active, as most of the time they do not trade as shown in the last row where we report the percentage of weeks over the total number of weeks of the trading period with open positions. For instance, the cotton pair opens its position only 2% of the weeks in our sample, while for natural rubber, it is 5%.

As a portfolio, however, the EVM pairs trading strategy is utilized 100% of the time, i.e., there is at least one commodity pair open in every week of the trading period. Overall, Table 3 shows that there are statistical arbitrage opportunities between the Chinese and International commodity futures pairs before transaction costs. The EVM strategy is actively utilized and it offers a plausible risk-return profile.

INSERT TABLE 3 HERE

We compare the profitability of the EVM pairs trading strategy with other traditional strategies in commodity futures markets. In particular, we compare the equal-weighted and monthly rebalanced long-only commodity futures portfolio and the traditional long-short term structure, momentum and value commodity futures portfolios. The *term structure* (TS, hereafter) portfolio builds on the theory of storage (Kaldor, 1939; Working, 1949; Brennan, 1958) which relates the slope of the term structure of commodity futures prices to inventory levels and to the costs/benefits of owning the physical commodity. A premium can be earned by taking long (short) positions in futures with high (low) roll-yields at the last week of the the formation period (*R*). The *momentum* portfolio (Miffre and Rallis, 2007) essentially buys past winners and sells past losers based on past performance over the previous year (or 52 weeks). Finally, the *value* portfolio (Asness et al., 2013) in commodity futures is based on long-term mean reversion. Specifically, value buys long-term losing (high-value or cheap) commodities and sells long-term winning (low-value or expensive) commodities based on the log of the average normalized commodity prices five years ago minus the log of the normalized price on the last week of the the formation period (*R*).

For consistency, we consider the conservative case with committed capital in the portfolio returns, reinvest the capital within the month and rebalance monthly in a fully-collateralized manner. We first split the cross-section of commodities (N=20) based on the median of the above signals (TS, momentum and value). We take a long position in the commodities which signal is above these median and a short position in the commodities which signal is below these median. Each of these positions is equally-weighted.[10]

Table 4 reports the performance for the various traditional trading strategies. Looking at the mean returns, we observe that neither the EW long-only nor the traditional long-short portfolios offer higher mean returns than the EVM pairs trading strategies. Their annual Sharpe ratios ranges from -0.18 (value portfolio) to 0.31 (TS portfolio), with very large maximum drawdowns ranging from -0.10 (TS portfolio) to -0.42 (EW portfolio). Except for the TS portfolio, all the CERs are negative.

---

[10] We may obtain a better performance in term structure, momentum, and value strategies if we trade in the extreme portfolios (e.g. the extreme quintiles), but for consistence with the pairs trading strategies, we use the median so we can cover all the N commodity futures.

INSERT TABLE 4 HERE

Figure 3 plots the monthly evolution of $1 invested at the beginning of the sample period in February 2009 for the long-only and long-short portfolios. A dollar invested in the EVM pairs trading strategies in February 2009 would be worth $1.22 in February 2018 which is superior to the traditional long-short portfolios with earnings of $1.14 (TS), $1.04 (Momentum) and $0.93 (Value), and comparable to long-only EW portfolio ($1.24). Moreover, the EVM pairs trading strategy offers more stable earnings than all the traditional strategies. The plot confirms the results of Tables 3 and 4, i.e. the EVM pairs trading strategy offers a stable and positive mean return over the sample period.

INSERT FIGURE 3 HERE

Table 5 reports the risk-adjusted performance (alpha) for the pairs trading strategies by sector and for each commodity pair. The benchmark model controls for the EW long-only portfolio, and the long-short TS, momentum and value risk premia. The pairs trading strategies offers a sizable alpha after accounting for the traditional portfolios. This sizable alpha is also observed in metal futures but not in agricultural futures. The only significant risk factor is the momentum factor which loads negatively in the pairs trading strategies. The value risk factor loads positively, albeit insignificant. These results are somehow expected as the pairs trading strategy exploits mean reversion in the relative prices.

INSERT TABLE 5 HERE

5. Robustness tests

To test the robustness of our results, we analyze the impact of transaction costs as well as the potential to data snooping issues.

5.1 Turnover and breakeven transaction costs

We measure the portfolio *turnover* (TO) to estimate how intensive the trading strategies are. As the pairs trading strategies (may) change positions within the trading period and are rebalanced to equal weights at the end of the month, we need to measure how many times the positions are open during that month. Transaction costs are only incurred when the positions are open, i.e. $s_{j,t} = \{-1, +1\}$. We take into account all this in the weekly *turnover* as follows,

$$Turnover_{P,t} = \sum_{i=1}^{N} \left(\left| w_{P,i,t+1} - w_{P,i,t^+} \right| \right); t = 1, \dots, T-1 \qquad (5)$$

where $w_{P,i,t}$ is the allocation to the ith asset at week t in the portfolio and $w_{P,i,t^+} \equiv w_{P,i,t} \times e^{s_{i,t} \cdot r_{i,t+1}}$ is the actual portfolio weight at the end of week $t+1$, where $r_{i,t+1}$ denotes the realized weekly excess return of the ith commodity futures contract from week t to $t+1$. Thus, the above *turnover* measure captures the mechanical evolution of the commodity futures contracts allocation weights due to within-week price dynamics.

Panel A of Figure 4 plots the average weekly *turnover* of the long-short trading strategies.[11] Given the conservative committed capital approach that we follow, all strategies are not trade intensive. In particular, *turnover* ranges between zero (e.g. a buy and hold strategy) and two (when all portfolio constituents change positions); therefore, these figures are very low. For instance, the EVM pairs trading strategy has an average *turnover* of 13.17% per week. The EVM pairs trading strategy is more trade intensive compared with traditional long-short, TS, momentum, and value portfolios with *turnover* of 12.46%, 7.32% and 4.43% per week, respectively.

INSERT FIGURE 4 HERE

The key question is whether *transaction costs* (*TC*) eliminate the performance of the pairs trading strategies. To get a sense of the effect of transaction costs in the global commodity futures markets, we measure the breakeven transaction costs as follows

$$\tilde{r}_{P,t+1} = r_{P,t+1} - TC \cdot Turnover_{P,t} \tag{6}$$

where $r_{p,t}$ is the portfolio return in week t, and $Turnover_{P,t}$ is the portfolio turnover from Equation (5). TC in Equation (6) represents the breakeven transaction costs, i.e. the round-trip transaction costs that makes the mean return of the portfolio equal zero.

Panel B of Figure 4 shows that the EVM pairs trading strategy is still profitable with *TC* lower than 0.32% per round trip. This breakeven level is higher than the transaction costs in

[11] For ease of comparison, we do not report the turnover and transaction costs of the long-only commodity futures portfolio because of as this strategy is passive, its turnover is not comparable with active long-short strategies.

commodity futures contracts.[12] Thus, it is clear that arbitrage opportunities in these commodity futures pairs will remain even after deducting conservative transaction costs. These breakeven transaction costs are higher than those of the traditional long-short portfolios with a maximum breakeven transaction cost of 0.22% (momentum portfolio).

5.2 Data snooping test

Repetitively employing the same dataset to assess the performance of many investment strategies can trigger false discoveries, a practice known as data snooping. To ensure our empirical results are free of data snooping, we conduct the Superior Predictive Ability (SPA) test of Hansen (2005). The idea is to create simulated scenarios with similar characteristics to the actual one and compare whether the simulated results are similar to the original ones. To do so, SPA compares the performance of a set of strategies in each simulated scenario and tests the null hypothesis whether the best strategy is not statistically superior to the benchmark which, in our context, is the EVM pairs trading strategy. Appendix A provides the technical details of the Hansen's SPA test.

As shown in Table 6, the bootstrap *p*-values of the test, ranging from 0.73 to 0.81 across all alternative functions, are distinctly large and unable to reject the null hypothesis, i.e., the best strategy is not statistically superior to the EVM pairs trading strategy. Thus, the profitable arbitrage opportunities between the Chinese and International commodity futures contracts is robust to data snooping bias.

INSERT TABLE 6 HERE

[12] For example, the commodity futures literature for the US markets documents transaction costs of 0.033% (Locke and Venkatesh, 1997) and 0.086% (Marshall et al., 2012) per round trip. Indriawan et al. (2018) document that the bid-ask spread for aluminum and copper futures in China are 0.040% and 0.024%, respectively.

6. Conclusions

We investigate the profitability of a pairs trading strategy using Chinese and International commodity futures during the period January 2004 to February 2018. We use a time-series approach where the commodity pairs have been pre-determined given similar underlyings. For a portfolio of commodities, this strategy yields an excess return of 2.08% p.a. with a Sharpe ratio of 0.79. For a portfolio of metal futures, this strategy yields 5.32% p.a. and a Sharpe ratio of 1.47. In the case of gold-only futures, this strategy yields 7.39% p.a. and a Sharpe ratio of 1.95.

This performance is not a compensation for exposure to traditional strategies based on term structure, momentum and value portfolios. Arbitrage opportunities in these commodity pairs remain even after accounting for transaction costs and are robust to data snooping bias. The profitability of the pairs trading strategy, especially for metal futures, is encouraging as the pairs trading literature (such as US stocks) documents declining trading profits over the decades (Do and Faff, 2010, 2012). Our results show that even in recent periods, a profitable pairs trading strategy in global commodity futures is still achievable. We also find that the more regulated commodities (e.g., wheat) perform worse in pairs trading. However, as China is gradually opening up its futures markets to foreign participation, pairs trading strategies would be more profitable.

Appendix A. Superior Predictive Ability test of Hansen (2005)

Using the EVM pairs trading strategy as benchmark, we compare the relative performance of the four traditional strategies, i.e. the long-only EW, and the long-short TS, momentum, and value portfolios. Let $r_{m,t}$ denote the week *t* excess returns of the *m*th strategy (with $m = 1, \dots, 5$) and define $r_t^{max} \equiv max(r_{m,t})$. Performance is measured in terms of the expected "loss" modelled as in Hansen (2005) with a linear mathematical function, $L_{m,t} \equiv r_t^{max} - r_{m,t}$, and two nonlinear functions, $L_{m,t} \equiv 1/exp(\lambda r_{m,t})$ with curvature parameter $\lambda = \{1, 2\}$. Specifically, the loss functions for the benchmark is defined as $L_{EVM,t} \equiv r_t^{max} - r_{EVM,t}$ and $L_{EVM,t} \equiv 1/exp(\lambda r_{EVM,t})$, respectively. The expected "loss" of the *m*th strategy relative to the benchmark is therefore $E[d_{m,t}] = E[\, L_{EVM,t} - L_{m,t}]$ for $t = 1, \dots, T$ weeks. Strategy m is better than the benchmark (EVM pairs trading strategy) if and only if $E[d_{m,t}] > 0$. The null hypothesis is that the best of the *M* strategies incurs a larger "loss" than the benchmark EWI strategy; *i.e.*, H_0: $E[d_{m,t}] \leq 0$, for all $m = 1, \dots, M$. In other words, we can interpret the null hypothesis of SPA test as the best strategy is not statistically superior to the EVM pairs trading portfolio.

The test is based on a statistic with a non-standard distribution which we approximate using the Politis and Romano (1994) bootstrap. For each strategy m, we construct B=10,000 time-series $\{d_{m,t}^*\}$ by combining randomly-sampled blocks of length l from the original time-series $\{d_{m,t}\}$. The block-length l is a geometrically distributed random variable with expected value $1/q$; we consider $q = \{0.2, 0.5\}$.

Reference List

Brennan, M. (1958). The supply of storage. *American Economic Review, 48*, 50-72.

Chen, D., Cui, J., Gao, Y., & Wu, L. (2017). Pairs trading in Chinese commodity futures markets: an adaptive cointegration approach. *Accounting and Finance, 57*, 1237-1264.

Do, B., & Faff, R. (2010). Does simple pairs trading still work? *Financial Analysts Journal*, 66, 83-95.

Do, B., & Faff, R. (2012). Are pairs trading profits robust to trading costs? *Journal of Financial Research*, 35, 261-287.

Elliott, R. J., Hoek, J. V. D., & Malcolm, W. P. (2005). Pairs trading. *Quantitative Finance, 5*, 271-276.

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

Fan, J. H., & Zhang, T. (2018). Demystifying commodity futures in China. *Working Paper*.

Fernandez-Perez, A., Frijns, B., Fuertes, & A.-M., Miffre, J. (2018). The skewness of commodity futures returns. *Journal of Banking and Finance*, *86*, 143-158.

Fung, H.-G., Leung, W.K., & Xu, X. (2003). Information flows between the US and China commodity futures trading. *Review of Quantitative Finance and Accounting*, *21*, 267-285.

Fung, H.-G., Tse, Y., Yau, J., & Zhao, L. (2013). A leader of the world commodity futures markets in the making? The case of China's commodity futures. *International Review of Financial Analysis*, *27*, 103-114.

Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies, 19*, 797-827.

Galenko, A., Popova, E., & Popova, I. (2012). *Journal of Alternative Investments*, *15*, 85-97.

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

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics, 23*, 365-380.

Indriawan, I., Liu, Q., & Tse, Y. (2018). Market quality and the connectedness of steel rebar and other industrial metal futures in China. *Working Paper*.

Kaldor, N. (1939). Speculation and economic stability. *Review of Economic Studies*, *7*, 1-27.

Koijen, R., Moskowitz, T., Pedersen, L. & Vrugt, E. (2018). Carry. *Journal of Financial Economics*, *127*, 197-225.

Krauss, C. (2017). Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys, 31*, 513-545.

Liu, Q., Tse, Y., & Zhang, L. (2018). Including commodity futures in asset allocation in China. *Quantitative Finance*, *18*, 1487-1499.

Locke, P., and Venkatesh, P. (1997) Futures market transaction costs. *Journal of Futures Markets*, *17*, 229-245.

Marshall, B. R., Nguyen, N. H., & Visaltanachoti, N. (2012). Commodity liquidity measurement and transaction costs. *Review of Financial Studies*, *25*, 599–638.

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

Patton, D. (2017). China cuts minimum purchase price for wheat for first time in over a decade. Reuters, October 27.

Politis, D.N., & Romano, J.P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*, 1303–1313.

Rad, H., Low, R.K.Y. and Faff, R.W. (2016). The profitability of pairs trading strategies: distance, cointegration, and copula methods. *Quantitative Finance*, *16*, 1541-1558.

Yang, Y., Goncu, A., & Pantelous, A. (2017). Pairs trading with commodity futures: Evidence from the Chinese market. *China Finance Review International, 7*, 274-294.

Working, H. (1949). The theory of price of storage. *American Economic Review*, *39*, 1254-1262.

Figure 1. Correlations between the commodity pairs.

This figure plots the pairwise correlations between the excess returns of the commodity pairs for the full sample from January 2004 to February 2018.

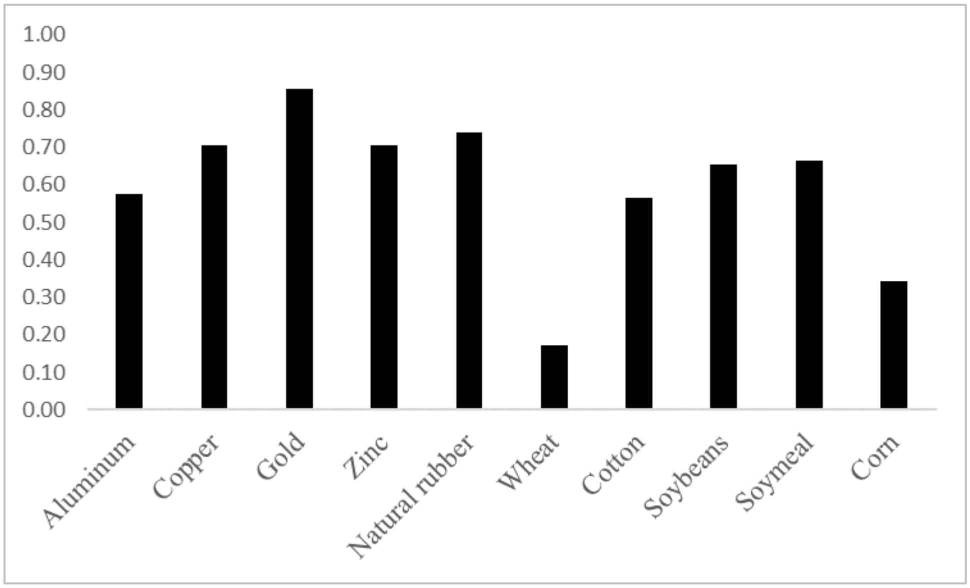


Figure 2. Weekly normalized prices: Aluminum.

This figure plots the normalized prices (cumulative log returns) for aluminum for the period February 2009 to February 2018. The left axis represents prices while the right-axis represents an indicator variable when the pairs trading strategy is triggered or not.

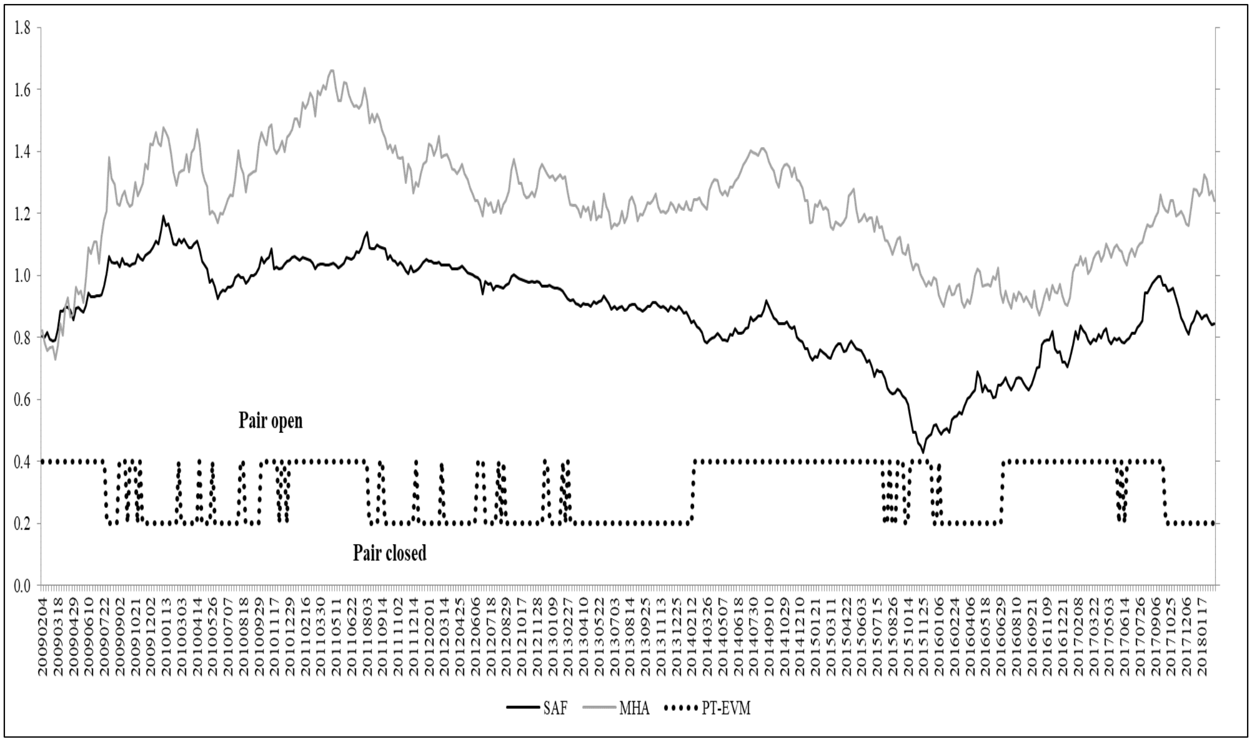


Figure 3. Future value of $1 invested in the trading strategies.

This figure plots the evolution of $1 invested in January 2009 using various trading strategies.

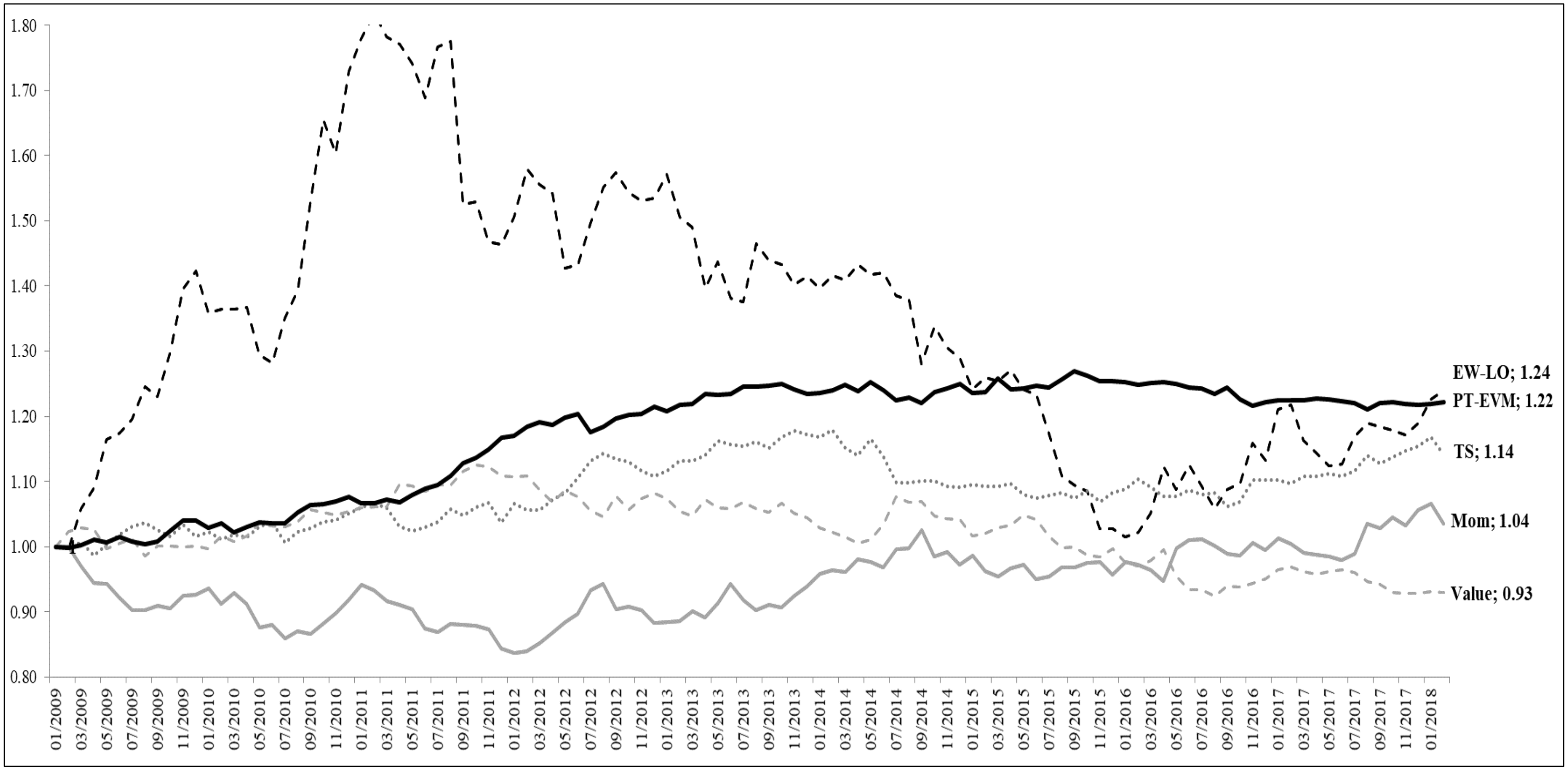
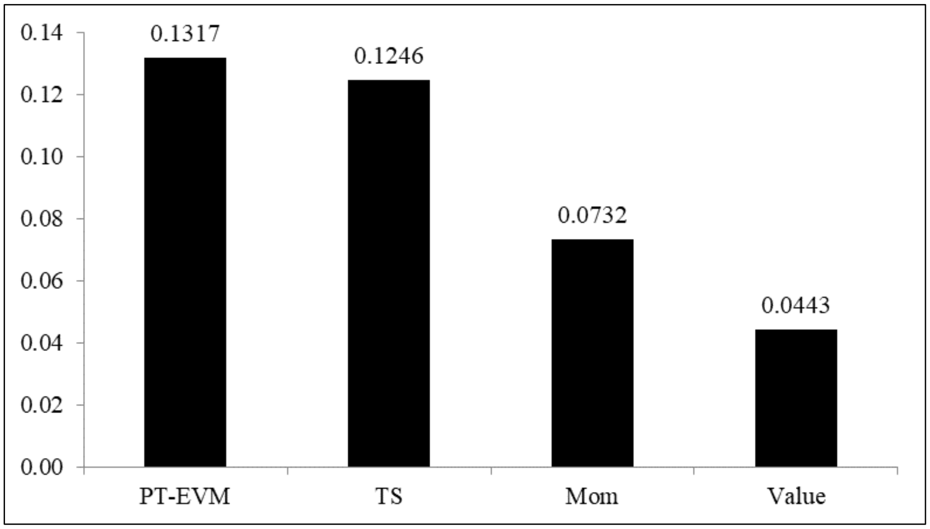


Figure 4. Turnover and breakeven transaction costs

This figure plots the turnover (Panel A) and the breakeven transaction costs (Panel B) of long-short trading strategies (EVM, TS, momentum and value portfolios).

Panel A: Turnover



Panel B: Breakeven transaction costs

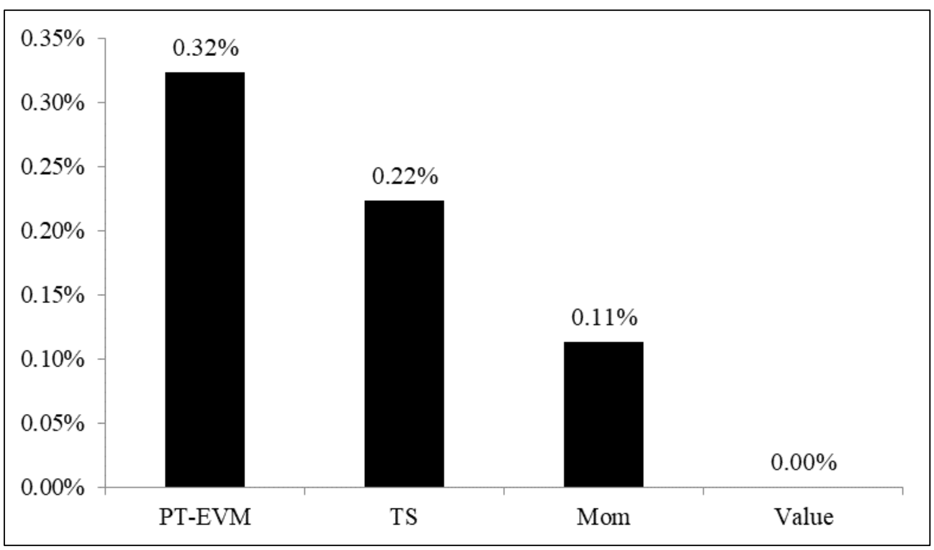


Table 1. Statistics for the Chinese and International commodity futures contracts

The table reports statistics for Chinese and International commodity futures contracts from January 2004 to February 2018. Mean and standard deviation (StDev) are annualized. Newey-West robust *t*-statistics are shown in parenthesis.

Commodity	Symbol	Exchange	Mean Returns	t-stat	StDev	Skewness	Kurtosis	Sharpe ratio
Chinese Commodity Futures								
Aluminum	SAF	Shanghai Futures Exchange	-0.011	(-0.24)	0.15	-0.35	3.98	-0.07
Copper	SCF	Shanghai Futures Exchange	0.128	(1.56)	0.26	-0.41	2.60	0.50
Gold	SHG	Shanghai Futures Exchange	0.025	(0.47)	0.18	-0.52	4.55	0.14
Zinc	ZNS	Shanghai Futures Exchange	-0.030	(-0.37)	0.24	-0.81	3.57	-0.12
Natural rubber	SNR	Shanghai Futures Exchange	-0.056	(-0.61)	0.30	-0.33	1.41	-0.19
Wheat	CWS	Zhengzhou Commodity Exchange	-0.045	(-1.36)	0.11	-0.67	4.93	-0.39
Cotton	CCF	Zhengzhou Commodity Exchange	-0.011	(-0.20)	0.17	-0.22	8.51	-0.06
Soybeans	DSA	Dalian Commodity Exchange	0.018	(0.39)	0.18	-0.58	3.30	0.10
Soymeal	DSM	Dalian Commodity Exchange	0.072	(1.23)	0.21	-0.35	2.13	0.35
Corn	DCC	Dalian Commodity Exchange	0.030	(0.90)	0.11	-0.04	1.61	0.28
International Commodity Futures								
Aluminum	MHA	LME (UK)	0.017	(0.24)	0.27	-0.09	2.28	0.07
Copper	MCU	LME (UK)	0.061	(0.63)	0.32	-0.93	6.83	0.19
Gold	GC	NYMEX (US)	0.065	(1.42)	0.19	-0.55	2.99	0.34
Zinc	MZS	LME (UK)	0.068	(0.73)	0.35	-0.62	3.35	0.20
Natural rubber	JRU	TCE (Japan)	-0.081	(-0.92)	0.32	-0.56	2.63	-0.26
Wheat	KW	KCBT (US)	-0.077	(-0.95)	0.30	0.40	1.65	-0.25
Cotton	CT	NYCE (US)	-0.042	(-0.53)	0.27	-0.13	0.84	-0.15
Soybeans	S	CBOT (US)	0.045	(0.66)	0.25	-0.54	2.63	0.18
Soymeal	SM	CBOT (US)	0.105	(1.34)	0.29	-0.20	2.02	0.37
Corn	C	CBOT (US)	-0.052	(-0.62)	0.30	-0.38	3.52	-0.17

Table 2. State-space parameters

This table reports the state-space parameters of Equations (1) and (2) for each pair of commodities over the full sample (from January 2004 to February 2018). Newey-West t-statistics are in parenthesis.

	Aluminum	Copper	Gold	Zinc	Natural rubber	Wheat	Cotton	Soybeans	Soymeal	Corn
a	-0.009***	0.003**	-0.009	-0.021*	0.003*	-0.001	0.014***	-0.002*	-0.004*	0.006
	(-3.00)	(2.02)	(-1.35)	(-1.71)	(1.76)	(-0.49)	(3.36)	(-1.66)	(-1.88)	(1.52)
b	0.974***	0.997***	0.986***	0.985***	0.994***	0.994***	0.973***	0.986***	0.990***	0.991***
	(118.21)	(428.52)	(95.60)	(112.89)	(289.45)	(186.59)	(114.80)	(137.88)	(197.74)	(163.67)
σ^2	0.0007***	0.0003***	0.0000***	0.0004***	0.0006***	0.0017***	0.0009***	0.0006***	0.0006***	0.0014***
	(8.60)	(7.48)	(2.58)	(5.78)	(7.01)	(9.22)	(7.61)	(10.48)	(6.19)	(6.53)
d^2	0.0001**	0.0004***	0.0001***	0.0004***	0.0002***	0.0000	0.0000	0.0000	0.0001***	0.0001
	(2.60)	(3.95)	(5.34)	(5.14)	(4.76)	(0.37)	(0.67)	(1.48)	(3.05)	(0.85)
ρ	0.026	0.003	0.014	0.015	0.007	0.006	0.027	0.014	0.010	0.009
μ	-0.34	0.91	-0.63	-1.36	0.44	-0.08	0.53	-0.14	-0.38	0.67

Table 3. EVM pairs trading

This table reports statistics for monthly excess returns of fully-collateralized portfolios from February 2009 to February 2018 for each pairs of commodities, for the metals and agricultural ("Agric.") portfolios, and also the portfolio with all the pairs ("All"). CER is annualized certainty-equivalent return with power utility preferences (CRRA parameter $\gamma = 5$). Mean and standard deviation (StDev) are annualized. Newey-West robust t-statistics are shown in parentheses. The last row reports the percentage of weeks with open positions according to each pairs trading strategy. ***, ** and * denote statistical significance at 1%, 5% and 10% levels, respectively.

	All	Metals	Agric.	Aluminum	Copper	Gold	Zinc	Natural rubber	Wheat	Cotton	Soybeans	Soymeal	Corn
Mean	0.021**	0.053***	-0.001	0.045*	0.037*	0.074***	0.059**	-0.009	-0.018	-0.001	0.005	0.007	0.013
	(2.20)	(3.40)	(-0.08)	(1.90)	(1.86)	(4.59)	(2.44)	(-0.75)	(-0.57)	(-0.16)	(0.25)	(0.70)	(0.33)
StDev	0.026	0.036	0.035	0.067	0.042	0.038	0.066	0.029	0.106	0.025	0.071	0.030	0.124
Downside volatility (0%)	0.017	0.025	0.026	0.048	0.014	0.010	0.039	0.025	0.077	0.021	0.052	0.020	0.096
Skewness	-0.26	0.214	-0.610***	-0.19	3.36***	2.04***	0.55**	-3.86***	-0.02	-3.56***	-0.16	0.92**	-0.56**
	(-1.09)	(0.91)	(-2.60)	(-0.82)	(14.32)	(8.68)	(2.36)	(-16.45)	(-0.09)	(-15.18)	(-0.67)	(3.91)	(-2.37)
Excess Kurtosis	0.19	1.83***	1.52***	3.55***	15.71***	5.44***	1.28***	40.75***	1.55***	56.60***	0.96**	13.25***	0.93**
	(0.41)	(3.90)	(3.25)	(7.56)	(33.48)	(11.60)	(2.73)	(86.84)	(3.31)	(120.63)	(2.05)	(28.24)	(1.98)
JB normality test p-value	0.429	0.006	0.005	0.001	0.001	0.001	0.009	0.001	0.013	0.001	0.068	0.001	0.018
99% VaR (Cornish-Fisher)	0.018	0.023	0.031	0.060	-0.012	0.000	0.035	0.076	0.084	0.097	0.054	0.037	0.101
% of positive months	59%	72%	50%	39%	23%	68%	50%	2%	37%	2%	37%	11%	55%
Maximum drawdown	-0.05	-0.05	-0.12	-0.11	-0.04	-0.02	-0.10	-0.05	-0.35	0.00	-0.15	-0.06	-0.30
Sharpe ratio	0.79	1.47	-0.02	0.67	0.88	1.96	0.89	-0.29	-0.17	-0.05	0.07	0.22	0.10
Sortino ratio	1.25	2.17	-0.03	0.94	2.71	7.62	1.53	-0.34	-0.24	-0.07	0.10	0.33	0.13
Omega ratio	1.79	3.36	0.98	1.88	4.34	8.14	2.13	0.36	0.87	0.80	1.07	1.48	1.08
CER	0.019	0.050	-0.004	0.034	0.032	0.070	0.048	-0.011	-0.047	-0.003	-0.007	0.004	-0.028
% of weeks with open position	100%	95%	100%	52%	25%	78%	70%	5%	77%	2%	60%	10%	88%

Table 4. Traditional strategies

The table reports statistics for monthly excess returns of fully-collateralized portfolios from February 2009 to February 2018. EW represents the equally-weighted long-only portfolio. TS represents the long-short term structure strategy. CER is annualized certainty-equivalent return with power utility preferences (CRRA parameter $\gamma = 5$). Mean and standard deviation (StDev) are annualized. Newey-West robust *t*-statistics are shown in parenthesis.

	EW	TS	Momentum	Value
Mean	0.030	0.014	0.004	-0.009
	(0.62)	(1.10)	(0.19)	(-0.58)
StDev	0.129	0.045	0.064	0.047
Downside volatility (0%)	0.085	0.031	0.039	0.029
Skewness	-0.18	-0.45*	0.02	0.15
	(-0.78)	(-1.92)	(0.09)	(0.64)
Excess Kurtosis	1.29***	0.23	-0.11	0.46
	(2.75)	(0.49)	(-0.23)	(0.99)
JB normality test p-value	0.024	0.091	0.500	0.420
99% VaR (Cornish-Fisher)	0.10	0.03	0.04	0.03
% of positive months	50%	57%	50%	44%
Maximum drawdown	-0.421	-0.099	-0.161	-0.178
Sharpe ratio	0.235	0.309	0.065	-0.183
Sortino ratio	0.359	0.444	0.108	-0.296
Omega ratio	1.199	1.252	1.048	0.872
CER	-0.012	0.009	-0.006	-0.014

Table 5. Risk-adjusted performance

The table reports the annualized alpha of the pairs trading strategies. EW represents the equally-weighted long-only portfolio. TS represents the long-short term structure strategy. The last row reports the adjusted-R^2. Newey-West robust *t*-statistics are shown in parenthesis. ***, ** and * denote statistical significance at 1%, 5% and 10% levels, respectively.

	All	Metals	Agric.	Aluminum	Copper	Gold	Zinc	Natural rubber	Wheat	Cotton	Soybeans	Soymeal	Corn
EVM alpha	0.022**	0.053***	0.001	0.044**	0.035*	0.073***	0.061***	-0.008	-0.018	-0.001	0.011	0.008	0.018
	(2.42)	(3.45)	(0.13)	(1.99)	(1.81)	(4.61)	(2.55)	(-0.73)	(-0.52)	(-0.08)	(0.56)	(0.85)	(0.44)
EW	-0.014	-0.021	-0.009	-0.023	-0.006	0.020	-0.075	-0.013	0.065	0.004	-0.017	-0.043	-0.044
	(-0.53)	(-0.88)	(-0.26)	(-0.60)	(-0.16)	(0.91)	(-1.36)	(-1.55)	(0.85)	(1.16)	(-0.32)	(-1.27)	(-0.37)
TS	0.030	0.106	-0.021	0.251*	0.095	0.095	-0.024	-0.065	0.094	-0.065	-0.138	0.070	-0.013
	(0.46)	(1.09)	(-0.25)	(1.88)	(0.84)	(0.85)	(-0.13)	(-1.30)	(0.34)	(-1.23)	(-0.91)	(1.27)	(-0.03)
Momentum	-0.157***	-0.070	-0.215***	-0.137	-0.021	-0.024	-0.097	0.038	-0.488**	0.002	-0.174	0.034	-0.716**
	(-3.39)	(-1.52)	(-2.90)	(-1.35)	(-0.46)	(-0.47)	(-0.86)	(1.03)	(-2.51)	(0.09)	(-1.30)	(0.90)	(-2.46)
Value	0.059	-0.001	0.099	0.092	-0.010	0.088	-0.175	-0.094	0.192	-0.015	0.300	0.136	0.053
	(0.98)	(-0.01)	(1.08)	(0.69)	(-0.13)	(0.98)	(-1.22)	(-1.10)	(0.66)	(-0.15)	(1.50)	(1.42)	(0.15)
Adj-R2	0.14	-0.01	0.18	0.00	-0.03	-0.01	0.00	0.00	0.08	-0.03	0.08	0.04	0.11

Table 6. Data-snooping-robust test for superior performance.

The table reports bootstrap *p*-values for the Superior Predictive Ability test of Hansen (2005) to control for data snooping bias in the comparison among the traditional long-short portfolios (EW, TS, Momentum, and Value) and the benchmark (the pairs trading strategy). The empirical distribution of the *t*-statistic is constructed by a random-length block bootstrap simulation with B=10,000 replications. The null hypothesis is that the best of the M portfolios is not statistically superior to the benchmark pairs trading portfolio.

Benchmark	**Loss function**	**Expected block length 1/q**	
		q=0.2	*q=0.5*
EVM	Linear	0.73	0.74
	Exp(λ = 1)	0.78	0.80
	Exp(λ = 2)	0.80	0.81