Journal of Banking & Finance 34 (2010) 409–426

Contents lists available at ScienceDirect

Journal of Banking & Finance

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

Trend-following trading strategies in commodity futures: A re-examination

Andrew C. Szakmary [a,*], Qian Shen [b], Subhash C. Sharma [c]

[a] Department of Finance, Robins School of Business, University of Richmond, Richmond, VA 23173, USA
[b] Department of Economics and Finance, School of Business, Alabama A&M University, Normal, AL 35762, USA
[c] Department of Economics, Southern Illinois University at Carbondale, Carbondale, IL 62901, USA

ARTICLE INFO

Article history:
Received 20 February 2008
Accepted 7 August 2009
Available online 11 August 2009

JEL classification:
G11
G13
G14

Keywords:
Trend-following
Trading rules
Momentum
Commodity futures

ABSTRACT

This paper examines the performance of trend-following trading strategies in commodity futures markets using a monthly dataset spanning 48 years and 28 markets. We find that all parameterizations of the dual moving average crossover and channel strategies that we implement yield positive mean excess returns net of transactions costs in at least 22 of the 28 markets. When we pool our results across markets, we show that all of the trading rules earn hugely significant positive returns that prevail over most subperiods of the data as well. These results are robust with respect to the set of commodities the trading rules are implemented with, distributional assumptions, data-mining adjustments and transactions costs, and help resolve divergent evidence in the extant literature regarding the performance of momentum and pure trend-following strategies that is otherwise difficult to explain.

1. Introduction

A momentum strategy is a simple trading rule whereby one rank-orders past returns on the assets being investigated, then takes long positions in assets that performed relatively well (past winners) and short positions in assets that performed relatively poorly (past losers). Following a momentum strategy is a bet that past relative performance will continue into the future. A large body of empirical research documents that momentum strategies appear to earn significant abnormal returns in a wide variety of markets; see, for example, Jegadeesh and Titman (1993, 2001) and Conrad and Kaul (1998) for evidence regarding momentum profits in US stocks, Rouwenhorst (1998) for European stocks, Moskowitz and Grinblatt (1999) for industry portfolios and, finally, Chan et al. (2000) for evidence of momentum profits when the strategy is implemented with countrywide stock market indices.[1]

Stock market momentum studies are often criticized on the grounds that the profits generated may be illusory. Korajczyk and Sadka (2004) and Lesmond et al. (2004) both find that once the direct and indirect transactions costs are taken into account, it is doubtful that momentum strategies would have yielded abnormal returns prior to the recent decimalization of stock price quotes. These difficulties are exacerbated by Lesmond et al. (2004) finding that the majority of momentum returns are apparently generated by return continuation among poorly performing stocks. Selling such stocks short, as required by a momentum strategy, could prove difficult due to the uptick rule, the inability to gain access to short sale proceeds and the general difficulty associated with borrowing shares of small, illiquid stocks. Partly as a result of these frictions, attention has recently shifted to the examination of momentum strategies in futures markets, where, as discussed in detail by Shen et al. (2007) and Marshall et al. (2008), transactions costs are considerably lower and taking short positions is relatively easy. Another reason momentum strategies are being examined in commodity futures is that several recent studies indicate that commodity futures provide substantial diversification benefits, especially when they are actively managed. See, for example, Gorton and Rouwenhorst (2006) and Erb and Harvey (2006).

Most closely related to this study, Shen et al. (2007) and Miffre and Rallis (2007) show that momentum strategies in commodity futures earn impressive returns that are too large to be subsumed by the relatively low transactions costs prevailing in these markets.

* Corresponding author. Tel.: +1 804 289 8251; fax: +1 804 289 8878.
E-mail addresses: aszakmar@richmond.edu (A.C. Szakmary), qian.shen@aamu.edu (Q. Shen), sharma@siu.edu (S.C. Sharma).

[1] Because data-mining is an important consideration in evaluating the performance of trading rules, it is important to note that Conrad and Kaul (1998), as well as Grundy and Martin (2001) show that the momentum strategies work in all subperiods they examine, being consistently profitable in the US stock market since the 1920s. Asem (2009) finds that momentum profits, albeit somewhat lower, continue to be significantly positive when portfolios are constructed using only dividend-paying stocks.

0378-4266/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbankfin.2009.08.004

Paradoxically, however, the performance of other trend following trading rules in commodity futures (which we argue below should perform similarly to momentum strategies) has been found to be underwhelming at best, especially if one considers data-mining issues. See, for example, the exhaustive review of this literature by Park and Irwin (2007), and recent studies by Park and Irwin (2005) and Marshall et al. (2008). As an illustration, Park and Irwin (2007, Table 3) report that of nine modern (post 1988) studies examining the profitability of trading rules in commodity futures markets, six studies report positive results, two negative, and one mixed. However, even in the studies that they classify as having "positive" results, typically only a subset of trading rules shows any promise, and the findings could easily be explained by data-mining (Sullivan et al., 1999); indeed, Marshall et al. (2008) report that 14 out of 15 commodities they examine fail to generate statistically significant profits after adjustment is made for data-snooping bias.[2]

In this paper, we examine the profitability (after deducting reasonable transactions costs) arising from the implementation of trend-following strategies on a long-term, 28 commodity, monthly dataset similar to the one used in Shen et al. (2007). We focus on six parameterizations each of both a dual moving average crossover strategy and a channel strategy, without optimization, and find that every parameterization tested yields a positive mean net return in at least 22 of the 28 markets over our full sample period. When we aggregate our results across markets, we find that virtually all rules tested yield hugely significant profits over our full sample; indeed, the dual moving average crossover and channel strategies generally produce higher mean returns and Sharpe ratios than momentum strategies. When we break down our pooled returns over four subperiods of approximately equal length, and/or if we restrict pooling to those markets with relatively high trading volume, we find that our results remain largely intact, with surprising consistency, for all of these subperiods and aggregations, albeit (consistent with previous studies) there is some evidence of reduced profitability in the most recent 1996–2007 subperiod. Finally, we show that inferences obtained using Newey and West (1987) *t*-statistics remain largely unaffected if we use bootstrapped empirical distributions to test statistical significance, that the trading rule returns we report are unlikely to be due to data mining, and that their returns remain fairly impressive even if we make highly pessimistic assumptions regarding the level of transactions costs.

Our paper is primarily motivated by the divergent findings in the literature regarding momentum and pure trend-following trading rules in commodities. One possible explanation for this divergence is that momentum strategies have both a cross-sectional and a time-series component. Conrad and Kaul (1998) argue that cross-sectional variation in the mean returns of individual securities (presumably based on differing exposures to systematic risk factors) plays an important role in their profitability, and could account for the success of momentum strategies. However, more recent findings cast substantial doubt on the Conrad and Kaul hypothesis. If momentum profits were due primarily to cross-sectional differences in mean returns, then past winners (losers) should continue to be superior (inferior) performers indefinitely into the future. However, Jegadeesh and Titman (2001) find that momentum returns are positive only during the first twelve months after portfolio formation; if anything, returns beyond this 12-month horizon are negative. Both Shen et al. (2007) and Miffre and Rallis (2007) report similar findings in commodity futures markets. In a related line of inquiry, Chen and Hong (2002) and Jegadeesh and Titman (2002) formally decompose momentum profits, and find that the profitability of momentum strategies is mostly due to time-series dependence in realized returns rather than cross-sectional variation in expected returns.

We suspect that the divergent findings regarding momentum and pure trend-following trading strategies in the commodity futures literature stem from differences in *research design* rather than from any inherent superiority in the momentum decision rule per se. Two design issues, in particular, stand out. First, most momentum studies typically examine monthly data across many decades and implement strategies in which both formation and holding periods extend out to twelve months. Both Shen et al. (2007) and Miffre and Rallis (2007) follow this approach and report significant profitability for formation and holding periods extending out to 9–12 months, although in both studies profitability is most uniformly significant for one month holding periods. In contrast, virtually all studies of technical trading rules in commodity futures, including Park and Irwin (2005) and Marshall et al. (2008), use daily data, and involve much shorter holding periods and more frequent trading. Given that momentum has been shown to be an intermediate horizon phenomenon, we conjecture that the performance of trend-following rules may be very different if implemented over such horizons.

A second design issue involves aggregation across assets examined. Momentum studies typically include a large cross-section of stocks, some of which are implicitly purchased and some of which are sold short, thus ensuring that much of the idiosyncratic risk associated with holding positions in individual stocks is eliminated. Both Shen et al. (2007) and Miffre and Rallis (2007) duplicate this strategy in commodity futures by aggregating results across the numerous separate markets they examine. In contrast, most studies that have tested trend-following trading rules in commodities have examined only a few individual futures markets, and have not aggregated the returns across them. While Lukac et al. (1988) and Park and Irwin (2005) do report portfolio results, these studies examine only 12 futures markets, and three of these are foreign exchange or financial futures. We conjecture that expanding the analysis to a broader array of commodities may produce very different results.

In summary, the uniqueness of our study is that we apply the research design of a momentum study (in terms of examining intermediate horizons and aggregating results across a broad array of commodities) to the evaluation of a group of trend-following trading rules. The balance of this paper is organized as follows: Our dataset, methodology for constructing unit value indices, and procedures to ensure that we control for price limits and other microstructure issues are described in Section 2. Our basic trading rule tests and results for individual commodity futures markets are reported in Section 3, and tests and results for aggregations across these markets are reported in Section 4. We conduct robustness tests, with respect to distributional assumptions, data-mining issues and assumed transactions costs, in Section 5. Finally, Section 6 concludes the paper.

2. Data and construction of unit value indices

From the Commodity Research Bureau (CRB) historical data CD, we extract daily closing prices, trading volume and open interest for 28 futures markets where the underlying asset is a single deliverable commodity, as well as for cash-settled futures on the Goldman Sachs (GS) Commodity Index. In each case, we use the nearby

[2] Another issue is that many studies examining trading rules have included currencies, in which central bank intervention may play a role in impeding the adjustment of exchange rates to new information; once these results are stripped out, evidence for the profitability of technical trading rules is often weaker. For recent evidence on the substantial profitability of trend-following rules in currencies, see Harris and Yilmaz (2009); for evidence and discussion regarding whether intervention is responsible for trading rule profitability in foreign exchange markets, see Szakmary and Mathur (1997), LeBaron (1999) and Neely (2002).

contract, rolling over to the next contract on the last day of the month before contract expiration. To avoid distortions caused by contract rollovers, we ensure that percentage price changes are always calculated using data from the same contract; thus, on rollover days we extract prices for both the nearby and first-deferred contracts.[3] Once we have obtained daily returns series (adjusted for rollover) by commodity, we construct daily unit value indices from the daily returns.

For the purpose of constructing series used as *inputs* for generating trading signals for the various trend-following strategies, we convert the data for each commodity future to a monthly frequency by sampling the daily unit value indices on the last trading day of each calendar month (although, when constructing monthly volume and open interest measures, we average the daily volume and open interest values within each calendar month). However, monthly series used for the measurement of gross (i.e. before transactions costs) *holding period returns* are constructed differently. This is primarily because some of the contracts included in our study have price limits, whereby the maximum allowable price fluctuation per day is limited by the exchange. On days when price limits are reached, trading effectively shuts down, and traders must wait until the limits are no longer binding before trading activity resumes.

To ensure that the results we report are not an artifact of price limits and that our strategies are actually implementable, holding period returns are measured using a dataset from which price limit days are removed. Specifically, we identify price limit days for each commodity from return and trading volume data, and we create a second monthly unit value index for each commodity which is sampled on the *first day of each calendar month that is not a limit day*. Thus, throughout the study we use unadjusted monthly unit value indices (which are sampled on the last day of each month) in the formation period to generate trading signals, and the adjusted unit values to compute returns during the holding period, which starts the following month. This procedure ensures that there is at least a one-day lag between the generation of trading signals and the taking of positions, and that both entry and exit trades are delayed until price limits are no longer binding.

The 28 individual commodity futures included in our study are the same as those in Shen et al. (2007); they represent a broad cross-section of agricultural, industrial, precious metal and energy futures markets, and specifically exclude currencies and other financial futures. The futures in our study are traded on several different exchanges and have different start dates. Eight of the commodities included in our study begin on July 1, 1959. Nine begin sometime in the 1960s, six in the 1970s, and the remaining five commodities start in 1980 or later; the GS index futures begin in July 1992. The last trading day for each commodity is December 31, 2007. Table 1 lists the 28 individual commodity futures and the GS index futures, along with their ticker symbols, associated exchanges and start dates. In addition, Table 1 provides market information, summary statistics, and trading volume for each of the commodity futures included in the study. The market information consists of the number of units of the underlying commodity deliverable per futures contract, the tick size and the average price of the nearby contract over the entire sample period.[4] Summary statistics relate to the returns associated with continuously maintained long positions in nearby contracts, without adjustment for the transactions costs that would be incurred when rolling over contracts. To facilitate return calculations over longer horizons and for consistency with other studies, we define the monthly return as $\mathrm{LN}(\mathrm{UV}_t/\mathrm{UV}_{t-1})$, where LN denotes a natural logarithm and UV is the unit value index of a particular futures contract at the end-of-month t, constructed as previously described.[5] We report mean returns to long positions, standard deviation, skewness, excess kurtosis, and the Jarque and Bera (1987) test statistic for the null hypothesis that the return is normally distributed. Finally, we report average daily trading volume for two time periods: the entire available sample for each commodity, and January 1996–December 2007. The trading volume figures reported in Table 1 are not specific to the nearby contract, as this information is not reported by the CRB. Rather, the figures are in terms of the notional dollar value of contracts for all maturity months, defined as total number of contracts traded times nearby futures price times contract multiplier.

The results in Table 1 reveal three things of particular interest. First, there are substantial differences in the volatility of returns across the 28 commodities. While monthly return standard deviations range from 2.66% (for domestic sugar) to 14.69% (Natural Gas), there are few obvious patterns or linkages between seemingly related markets: for example, the standard deviation for world sugar is nearly five times as high as for domestic sugar, the standard deviation for both silver and platinum are substantially higher than for gold, and hog futures (live/lean hogs, pork bellies) appear to be considerably more volatile than cattle futures. Second, we find that none of the individual commodity futures exhibit normality in their monthly returns. Return skewness is positive in 25 of the 28 individual commodities, and all of the return series (except the GS index) have positive excess kurtosis, indicating that most of these markets, similar to other financial markets, have fat-tailed distributions. The Jarque-Bera test formally rejects the normality assumption at the one percent level or better for all individual commodities, but not the GS index. These results underscore the need to conduct bootstrap tests on trading rule returns in order to verify that conventional t-statistics (which assume normal distributions) yield correct inferences. Finally, we note that there are extremely large variations in average trading volume among these markets, ranging from over 4868 million dollars per day in crude oil all the way down to 8.27 million dollars daily in oats. The low trading volume in some of these markets suggests that the trading strategies we examine may be difficult to implement on a large scale. Consequently, following Shen et al. (2007), in portfolio tests, we also focus on a subset of these 28 markets that exhibit relatively high trading volume, where the trading strategies are more likely to be implementable.

3. Basic trading rule tests and results for individual commodities

In this study, our main focus is the profitability associated with dual moving average crossover (DMAC) and channel strategies implemented in our sample of 28 commodity futures markets, and how the profitability of these pure trend-following strategies compares with those arising from momentum strategies examined

[3] This issue does not arise for trading volume or open interest, because the CRB reports total volume and open interest across all contracts currently traded; i.e. volume and open interest are not specific to one expiration month.

[4] To provide consistency across markets, we convert the tick size and the average nearby contract price to dollars per unit of the underlying commodity in those cases where these are specified as cents per unit.

[5] Here and throughout the study, our methodology for calculating gross returns implicitly assumes zero leverage, i.e. that traders deposit the full notional contract value at the time they take a position into their margin account, and earn the risk-free rate on margin account balances. Thus the returns reported in our study are best interpreted as unlevered excess returns. We recognize that most futures traders use leverage and would earn considerably higher mean returns than those we report. However, because both the mean and standard deviation of returns is linearly related to leverage, the *t*-statistics and Sharpe ratios associated with trading strategies we examine would be unaffected by the degree of leverage employed.

Table 1
Market information and summary statistics.

Symbol	Commodity	Exchange	Start date	Tick size	Contract multiplier	Average price	Monthly excess returns on long positions					Avg. trading volume ($ millions)	
							Mean (%)	Std. dev. (%)	Skewness (%)	Kurtosis (%)	Jarque–Bera	Overall	1996–2007
BO	Soybean Oil	CBOT	7/1/1959	0.00010	60,000	0.1941	0.7747	9.05	138.43	523.39	848.74**	199.12	419.68
C–	Corn	CBOT	7/1/1959	0.00250	5000	2.2154	–0.2114	6.55	136.10	930.22	2274.00**	557.33	1324.16
CC	Cocoa	NYBOT	7/1/1959	1.00000	10	1394.1271	0.3181	9.05	87.47	249.23	224.46**	64.96	141.63
CL	Crude Oil	NYMEX	3/30/1983	0.01000	1000	28.2145	1.2856	9.26	44.26	232.56	76.63**	4868.06	8717.77
CT	Cotton	NYBOT	7/1/1959	0.00010[a]	50,000	0.5471	0.0386	6.09	12.14	124.29	38.82**	188.89	389.70
FC	Feeder Cattle	CME	11/30/1971	0.00025	50,000	0.7305	0.3118	5.06	–31.27	280.61	149.12**	88.76	144.90
GC	Gold	NYMEX	12/31/1974	0.10000	100	366.0342	0.0402	5.41	75.72	753.96	975.81**	1510.20	2129.69
HG	Copper High Grade	NYMEX	11/1/1969	0.00050	25,000	0.9779	0.5959	7.92	56.76	328.63	229.67**	231.49	423.74
HO	Heating Oil	NYMEX	11/14/1978	0.00010	42,000	0.7908	1.1189	9.34	89.21	400.61	279.67**	1105.88	2089.47
HU/RB	Gasoline	NYMEX	12/3/1984	0.00010	42,000	0.8110	1.7386	10.29	97.04	331.17	169.44**	1311.14	2128.21
JO	Orange Juice	NYBOT	2/1/1967	0.00050	15,000	1.0001	0.5519	9.74	223.51	1292.57	3819.0**	26.77	49.42
KC	Coffee	NYBOT	8/1/1972	0.00050	37,500	1.1638	0.6807	11.30	134.52	384.04	389.34**	274.46	480.44
KW	Wheat, #2 Winter	KCBT	1/5/1970	0.00250	5000	3.5169	0.5152	7.23	206.73	1695.06	5771.25**	116.83	229.00
LB	Lumber	CME	10/1/1969	0.10000	110	223.2765	–0.1299	8.53	39.39	96.09	29.46**	31.13	34.00
LC	Live Cattle	CME	11/30/1964	0.00025	40,000	0.5961	0.5544	5.32	–1.62	252.06	136.89**	396.23	620.90
LH	Live/Lean hogs	CME	2/28/1966	0.00025	40,000	0.5709	0.6208	7.94	17.28	141.40	44.32**	197.43	326.46
NG	Natural Gas	NYMEX	4/4/1990	0.00100	10,000	3.7735	0.2776	14.69	56.49	46.71	13.20**	2587.01	3688.43
O–	Oats	CBOT	7/1/1959	0.00250	5000	1.3469	–0.1005	8.48	240.18	2162.62	11880.68**	8.27	13.81
PA	Palladium	NYMEX	1/3/1977	0.05000	100	196.2345	0.7340	11.35	264.18	2340.38	8898.64**	11.38	20.29
PB	Pork Bellies	CME	11/15/1964	0.00025	40,000	0.5935	0.3226	10.65	43.55	169.37	77.98**	93.50	32.61
PL	Platinum	NYMEX	3/4/1968	0.10000	50	437.8201	0.2155	8.21	–70.87	1196.55	2885.51**	51.03	54.63
RR	Rough Rice	CBOT	8/20/1986	0.00500	2000	7.8267	–0.3014	8.54	155.31	782.41	758.85**	9.25	12.93
S-	Soybeans	CBOT	7/1/1959	0.00250	5000	5.4000	0.4014	7.79	120.87	919.36	2187.60**	1259.08	2224.35
SB	Sugar #11/World	NYBOT	1/4/1961	0.00010	112,000	0.0923	0.5086	12.82	107.79	375.31	439.45**	182.59	400.88
SE	Sugar #14/Dom.	NYBOT	7/7/1987	0.00010	112,000	0.2165	0.0019	2.66	14.24	275.37	78.24**	12.66	13.11
SI	Silver	NYMEX	3/12/1967	0.00500	5000	6.0158	0.4354	10.67	298.41	3813.77	30360.89**	474.08	646.00
SM	Soybean Meal	CBOT	7/1/1959	0.10000	100	154.0248	0.7600	8.89	121.96	921.95	2201.73**	268.44	597.27
W–	Wheat, #2 Soft Red	CBOT	7/1/1959	0.00250	5000	3.0625	–0.0262	7.16	231.51	2201.77	12254.74**	287.01	687.76
GI	GS Commodity Index	CME	7/28/1992	0.05000	250	255.9574	0.5611	5.63	–5.99	–5.65	0.14	134.34	163.70

Notes: The exchange abbreviations are: CBOT = Chicago Board of Trade, CME = Chicago Mercantile Exchange, KCBT = Kansas City Board of Trade, NYBOT = New York Board of Trade, NYMEX = New York Mercantile Exchange. Returns reported in this table are gross returns; no allowance is made for transactions costs associated with rolling over contracts. The Jarque–Bera statistic tests the null hypothesis that the returns are normally distributed. Dollar trading volume is the notional value of contracts for all maturity months traded on an average day, defined as the total number of contracts traded × futures price × contract multiplier.

[a] The tick size in Cotton increases to $0.00050 per pound when the futures contract price exceeds $0.95 per pound.

* Indicate statistical significance at the 5% level.

** Indicate statistical significance at the 1% level.

by Shen et al. (2007) and Miffre and Rallis (2007).[6] Unlike these studies, ours focuses exclusively on 1-month holding periods. To measure gross returns associated with momentum strategies, we proceed as follows: At the end of each calendar month, we rank all eligible commodities independently on the basis of past total return to a long position, where the return for each commodity during the formation period is measured as the total log percentage change in the unit value index over the entire period. We use six different formation periods, i.e. 1, 2, 3, 6, 9 and 12 months. Based on each commodity's past relative return, we then assume a long position is taken in the commodity if it is a past "winner" (top 1/3 of commodities, based on formation period return), a short position if it is a past "loser" (bottom 1/3) and no position if that commodity ranks in the middle third.[7] We then measure 1-month holding period returns for each commodity, using a different unit value series which is sampled on the first trading day of each calendar month that is not a price limit-impacted day. Because the GS index is an agglomeration of numerous different commodities and momentum strategies are implemented via cross-sectional ranking of past returns, following earlier studies, we do not include the GS index futures in our momentum evaluations.

The net returns to momentum strategies are obtained by subtracting total assumed transactions costs each month, as a percent of contract value, from the gross returns, further assuming that one round-turn trade occurs each calendar month. We assume transactions costs consist of a fixed brokerage commission of $10 per contract (a representative fee that a discount broker might charge) and a bid-ask spread of one tick; numerous studies, e.g. Followill and Rodriguez (1991) and Locke and Venkatesh (1997) have estimated effective bid-ask spreads in commodity futures markets and have concluded that they are generally less than or equal to the value of one tick per contract. Thus, we estimate transaction cost as a percent of notional contract value in month t (TC_t) as:

$$TC_t = [10 + (\text{Tick Size } \times \text{CM})]/(\text{Price}_t \times \text{CM}), \quad (1)$$

where the tick size is measured in dollars, CM is the contract multiplier (i.e. the number of units of the underlying commodity deliverable per contract) and Price_t, again measured in dollars, is that of the nearby contract at the end-of-month t. We recognize that the above estimates of TC_t may be too low if, due to large scale trading activity, a trader temporarily influences the direction of market prices. Consequently, as part of our robustness tests in Section 5 below, we examine the impact of a more pessimistic transaction cost assumption, where the brokerage commission is assumed to be $20 per contract, and the bid-ask spread three ticks instead of one.

The mean net returns, by commodity, to all six parameterizations of the momentum strategies that we estimate are reported in Table 2. The t-statistics reported in this and subsequent tables are based on Newey and West (1987) standard errors with five lags.[8] Not surprisingly, in light of the results reported in Shen et al. (2007), we find that these strategies tend to earn positive net returns; the number of markets with positive mean returns ranges from 20 (with a 1-month formation period) to 25 when using a 3-month formation period (out of a total of 28 markets examined). If we assume return independence across markets, then the proportion of markets with positive returns is generally larger than can be explained purely by chance. The cumulative binomial probability function for X or fewer successes in N independent trials is given by:

$$\text{CP}(X) = \sum_{X=0}^{N} \frac{N!P^X(1-P)^{N-X}}{X!(N-X)!} \quad (2)$$

where P is the probability of success in each trial. We define "success" as a positive mean net return, and set $P = 0.5$ and $N = 28$. Because we do not know, a priori, if a preponderance of mean returns will be positive or negative, two-tailed probability values, calculated as $2[1 - \text{CP}(X)]$, are reported in Table 2, with X set to equal the observed number of markets with positive mean return minus 1. For the 3, 6 and 12-month formation period, we find that these binomial probabilities are less than 0.01, while for the remaining formation periods they are less than 0.05.[9] Another aspect of the results in Table 2 that deserves mention is that, regardless of formation period length, we never obtain statistically significant positive returns for more than six of the 28 commodities. Consequently, if, like most previous studies, ours did not examine such a broad array of markets, it is unlikely that we would be able to report conclusive results.

Net returns, by commodity, arising from six different parameterizations of a dual moving average crossover (DMAC) strategy are reported in Table 3, and net returns from six channel strategies in Table 4. In the DMAC strategy, a long position is taken in a commodity if the short-term moving average unit value (STMA) exceeds the long-term moving average unit value (LTMA) by B percent, i.e. if STMA > LTMA * (1 + B), and a short position is taken if STMA < LTMA * (1 − B). No position in a commodity is taken when STMA is within the band, i.e. when LTMA * (1 − B) < STMA < LTMA * (1 + B). Only month-end unit values are used in the above calculations. We consider an STMA of 1 or 2 months, and LTMA of 6 or 12 months. For B we use either 5% per annum, which implies that $B = 0.025$ and 0.05, respectively, at 6 and 12 month horizons for LTMA, or we estimate the DMAC with no band, i.e. $B = 0$. In the channel strategy, a long position is taken in a commodity if the latest end-of-month unit value exceeds the maximum of the end-of-month unit values over the previous L months, and a short position if the latest unit value is less than the minimum of the end-of month unit values over the previous L months. If the latest unit value is between the minimum and maximum observed over the previous L months, no position is taken in the commodity. We examine channel rules with lag lengths (L) of 3, 4, 5, 6, 9 and 12 months.[10]

The DMAC results in Table 3, and the channel results in Table 4, are similar – albeit somewhat more pronounced – than the momentum results in Table 2. Even the worst performing DMAC parame-

[6] Both DMAC and Channel strategies have been extensively used in previous studies. For example, Brock et al. (1992) and Sullivan et al. (1999) apply DMAC strategies to the S&P 500 index, and Szakmary and Mathur (1997) use DMAC rules in currencies. Channel strategies have been applied in currency futures by Taylor (1994) and in a broad cross section of futures by Lukac and Brorsen (1990). Both of these strategies are among the 12 examined in Lukac et al. (1988) and Park and Irwin (2005).

[7] Our sorting procedure follows Shen et al. (2007). Miffre and Rallis (2007) sort based on quintiles and obtain roughly similar results for momentum strategies.

[8] We use five lags because the commodities for which we have the most data have 482 observations, and (in the absence of overlapping data where the lag length is determined a priori) Newey and West (1987) recommend a lag length of $N^{0.25}$. Most inferences are not materially different using standard t-statistics.

[9] We must caution, however, that independence of returns across markets is likely not a valid assumption, in which case the binomial probability test understates the true probability of obtaining a given number of positive mean returns. Partly for this reason we later conduct portfolio tests, in which return independence is not assumed.

[10] Keeping in mind the pernicious effects of data mining, we made every possible attempt to *avoid* optimizing the parameters used in the DMAC and channel strategies. For the channel strategies, we report results for all the strategies we estimated. For the DMAC strategies, however, we were forced to try alternative parameterizations of the band width, because the initial band width chosen (20% per annum) resulted in too many commodities being assigned neutral, as opposed to long or short positions. In general, we found that higher values of B in the DMAC strategies result in higher mean returns, but because they are "out of the market" more often, the number of observations is fewer, and the incidence of statistically significant positive returns does not necessarily increase. We selected a band width of 5% per annum because this resulted in neutral positions being taken approximately 1/3 of the time, as in the momentum strategies.

Table 2
Net returns by commodity to momentum strategies.

Commodity	1-month FP		2-month FP		3-month FP		6-month FP		9-month FP		12-month FP	
	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat
BO	0.2184%	0.5092	0.8545%	1.9480	0.6703%	1.6646	0.8070%	1.9251	0.5248%	1.2179	0.4229%	0.9465
C	0.5837%	1.7179	0.5807%	1.8943	0.1685%	0.5105	−0.0773%	−0.2509	−0.1837%	−0.5305	−0.4452%	−1.1621
CC	−0.5074%	−1.1577	−0.2630%	−0.6372	0.2666%	0.6240	−0.4434%	−0.9848	0.2454%	0.5990	0.3553%	0.7965
CL	0.8752%	1.2872	0.3257%	0.4756	0.9427%	1.4198	0.5575%	0.9464	0.5373%	0.9260	0.7633%	1.1872
CT	0.2296%	0.7931	0.1799%	0.5271	0.7623%	2.3739*	0.3770%	1.1812	0.0701%	0.1990	0.0639%	0.1891
FC	0.1808%	0.7074	−0.0947%	−0.3054	0.0748%	0.2201	0.0830%	0.2009	0.3525%	0.9369	0.7207%	1.9925
GC	0.7533%	1.8210	1.0308%	2.5514*	0.6994%	1.5275	0.8724%	2.1371*	0.9152%	2.4711*	0.8015%	1.9918
HG	0.6257%	1.3460	0.8242%	1.6279	0.4760%	0.8634	0.7258%	1.3819	0.2888%	0.5213	0.1586%	0.2885
HO	1.3141%	2.7640**	0.4092%	0.8229	0.4458%	0.8781	−0.0334%	−0.0609	−0.0564%	−0.1059	0.4317%	0.7583
HU	0.5141%	0.6877	−0.6085%	−0.8680	−0.1846%	−0.2638	0.0082%	0.0127	0.6353%	1.0689	0.7617%	1.0596
JO	−0.4530%	−0.9804	0.1944%	0.4015	0.0381%	0.0791	0.0822%	0.1752	−0.3075%	−0.6153	−0.3250%	−0.7171
KC	−0.7615%	−1.1859	0.2985%	0.4277	0.3043%	0.4776	−0.2109%	−0.2983	0.5190%	0.7424	0.6337%	0.9643
KW	0.7243%	1.7126	0.5929%	1.5118	0.5752%	1.4379	0.0656%	0.1314	0.1587%	0.3304	1.3014%	2.4297*
LB	0.0197%	0.0376	0.3646%	0.7524	0.7169%	1.4725	0.3699%	0.6842	0.9533%	1.8746	1.2068%	2.3207*
LC	0.4689%	1.5807	0.2367%	0.7719	0.1598%	0.4953	0.5133%	1.6247	0.6476%	2.1409*	0.7162%	2.4645*
LH	−0.4273%	−1.0559	0.2692%	0.6452	0.3244%	0.8207	−0.0625%	−0.1420	0.3951%	0.9835	0.9390%	2.5822**
NG	1.4258%	1.5203	0.3587%	0.3316	0.6576%	0.5639	1.0809%	1.1620	0.1602%	0.1484	0.6379%	0.6390
O	0.0536%	0.1337	0.2675%	0.6970	0.0002%	0.0004	0.1989%	0.4737	0.1468%	0.3704	0.0210%	0.0520
PA	2.7586%	3.8164*	2.1171%	3.4851**	2.4916%	5.0445*	1.8096%	3.5397**	1.7693%	3.4498**	1.4285%	2.8780**
PB	−0.4553%	−0.8999	−0.0104%	−0.0187	0.1797%	0.3345	−0.2421%	−0.4342	−0.3952%	−0.6568	−0.0867%	−0.1474
PL	−0.0524%	−0.0935	−0.2174%	−0.4010	0.4873%	0.9760	0.3112%	0.5857	0.8176%	1.5282	0.6534%	1.1923
RR	0.5454%	0.8436	0.8699%	1.2903	1.0092%	1.3542	0.6150%	0.9390	−0.0945%	−0.1334	−0.2527%	−0.3259
S	−0.2741%	−0.6928	0.7381%	1.6077	−0.2145%	−0.4875	0.0231%	0.0543	−0.3359%	−0.7100	−0.9083%	−1.8012
SB	1.6638%	2.8169**	1.8718%	3.2051**	1.3542%	2.2678*	0.8995%	1.3948	0.1434%	0.2127	0.4673%	0.6985
SE	−0.1270%	−0.3836	−0.2367%	−0.6839	−0.2762%	−0.8174	0.1461%	0.5337	−0.0406%	−0.1776	−0.3147%	−1.6700
SI	0.0096%	0.0171	0.6562%	1.1516	0.2403%	0.4185	0.1199%	0.1934	0.5145%	0.9639	0.3903%	0.6990
SM	0.4461%	0.9805	0.5582%	1.4841	0.2694%	0.6904	0.4209%	0.9289	0.3360%	0.6792	0.1913%	0.3996
W	0.4344%	1.2540	−0.0939%	−0.2813	0.2951%	0.8240	0.0201%	0.0566	0.5532%	1.6160	0.9411%	2.5898**
Number of mkts. with positivemean return	20		21		25		22		21		22	
Binomial prob.	0.0357		0.0125		0.0000		0.0037		0.0125		0.0037	

Notes: All returns are unlevered monthly returns net of assumed transactions costs. The holding period for each strategy is assumed to be 1 month, i.e. formation period returns are updated at the end of each calendar month and a new position in each commodity (long, short or out of the market) is taken for the following month based on relative return ranking during the formation period. The reported t-statistics are for the two-tailed test that the mean net return differs from zero, and are calculated using Newey and West (1987) standard errors.

* Indicate statistical significance at the 5% level.

** Indicate statistical significance at the 1% level.

Table 3
Net returns by commodity to dual moving average crossover strategies.

Commodity	ST = 1, LT = 6, B = 0.025		ST = 1, LT = 12, B = 0.05		ST = 2, LT = 6, B = 0.025		ST = 2, LT = 12, B = 0.05		ST = 1, LT = 6, B = 0		ST = 1, LT = 12, B = 0	
	Mean	t-stat	Mean	t-stat	Mean	t-stat	Mean	t-stat	Mean	t-stat	Mean	t-stat
BO	0.8965%	2.8916**	0.6415%	1.9712*	0.7320%	2.2756*	0.5286%	1.6155	0.9890%	2.8468**	0.5063%	1.3533
C	0.5314%	2.3863*	0.3626%	1.7366	0.2765%	1.2371	−0.0114%	−0.0513	0.4007%	1.6758	0.2567%	1.0158
CC	0.1361%	0.3860	0.2983%	0.8269	0.1705%	0.4841	0.3858%	1.0881	0.1177%	0.3044	0.3601%	0.9498
CL	0.4335%	1.8991	0.3691%	1.6662	0.2529%	1.2080	0.2859%	1.2102	0.4533%	1.8048	0.3779%	1.4755
CT	0.4824%	2.1262*	0.3982%	1.7570	0.4182%	1.8951	0.2263%	1.0217	0.3941%	1.5066	0.3798%	1.5058
FC	0.2280%	1.5010	0.2252%	1.4938	0.2037%	1.3730	0.1787%	1.2724	0.2101%	1.2747	0.2668%	1.5409
GC	0.3442%	1.8071	0.3916%	2.4062*	0.2664%	1.5920	0.3943%	2.4764*	0.3668%	1.7849	0.5834%	3.3041**
HG	0.4850%	1.5621	0.4525%	1.5454	0.5092%	1.8094	0.3958%	1.3881	0.4757%	1.4050	0.5324%	1.6962
HO	0.3788%	1.7298	0.1717%	0.7429	0.0759%	0.3506	−0.0270%	−0.1118	0.3022%	1.2173	0.1461%	0.5647
HU	0.1298%	0.5664	0.2336%	0.9904	−0.0197%	−0.0839	0.1498%	0.6463	0.3546%	1.5264	0.2252%	0.9171
JO	0.2926%	1.0420	−0.0719%	−0.2469	0.2327%	0.8504	−0.1121%	−0.3754	0.1306%	0.4094	−0.0395%	−0.1294
KC	0.1775%	0.4573	0.1570%	0.3984	0.4048%	1.1826	0.3777%	0.9543	0.0040%	0.0095	0.0391%	0.0889
KW	0.4825%	2.1390*	0.5371%	2.2402*	0.3230%	1.6028	0.3748%	1.5187	0.5015%	2.0416*	0.5421%	2.0437*
LB	0.6411%	2.2985*	0.7284%	2.5002*	0.6041%	2.1135*	0.6174%	2.0915*	0.9118%	3.0711**	0.7686%	2.3364*
LC	0.1590%	0.9489	0.3332%	2.1803*	0.1416%	0.8367	0.3053%	1.8436	0.2251%	1.2119	0.3976%	2.0485*
LH	0.4104%	1.5449	0.5640%	2.2937*	0.5120%	2.0080*	0.4548%	1.8653	0.4347%	1.4949	0.4980%	1.7042
NG	0.4088%	1.2371	0.3196%	0.9900	0.2096%	0.5585	0.2041%	0.6422	0.3104%	0.8678	0.3121%	0.9051
O	−0.1377%	−0.4466	0.1001%	0.3558	−0.1448%	−0.4536	−0.0938%	−0.3293	−0.1950%	−0.5576	−0.0172%	−0.0528
PA	1.3139%	4.1212**	1.2745%	4.3313**	0.7680%	2.5487*	0.9264%	3.2792**	1.3528%	4.0459**	1.3725%	4.5347**
PB	−0.2315%	−0.6448	−0.2964%	−0.8598	−0.1257%	−0.3313	0.0438%	0.1198	−0.3853%	−0.9331	0.0331%	0.0808
PL	0.5642%	1.9043	0.4309%	1.4880	0.4906%	1.7152	0.3724%	1.3509	0.5722%	1.8694	0.4200%	1.3213
RR	0.5065%	2.1591*	0.2791%	1.2775	0.5352%	2.4848*	0.2576%	1.2045	0.5165%	2.1464*	0.4079%	1.8711
S	0.1317%	0.4963	0.1251%	0.4694	0.1590%	0.5980	−0.0098%	−0.0367	0.1725%	0.5874	0.1172%	0.4120
SB	1.6755%	3.6776**	1.3505%	2.7265**	1.1690%	2.4938*	1.0459%	2.1169*	1.5169%	3.1480**	1.3092%	2.4464*
SE	0.0396%	0.8333	0.0500%	1.0602	−0.0023%	−0.0409	−0.0227%	−0.4756	0.0870%	1.3109	0.0528%	0.8262
SI	0.2635%	0.7696	0.4310%	1.2588	0.4082%	1.1752	0.4957%	1.4093	0.1445%	0.4016	0.4925%	1.2724
SM	0.4015%	1.2667	0.5103%	1.6170	0.4593%	1.4354	0.3230%	0.9893	0.3874%	1.1392	0.3622%	1.0519
W	0.4295%	1.8117	0.5605%	2.1819*	0.2145%	0.9306	0.4746%	1.8792	0.4466%	1.6403	0.6332%	2.2521*
GI	0.2163%	1.8757	0.3091%	2.6541**	0.1569%	1.2981	0.2856%	2.4761*	0.3461%	2.6927**	0.3247%	2.4615*
Number of markets with positive mean return	26		26		24		22		26		26	
Binomial prob.	0.0000		0.0000		0.0002		0.0037		0.0000		0.0000	

Notes: All returns are unlevered monthly returns net of assumed transactions costs. The holding period for each strategy is assumed to be 1 month, i.e. position indicators are updated at the end of each calendar month and a new position in each commodity is taken for the following month. A long position is taken if the short-term moving average unit value (STMA) exceeds the long-term moving average unit value (LTMA) by B percent, i.e. if STMA > LTMA * (1 + B), and a short position is taken if STMA < LTMA * (1 − B). No position in a commodity is taken when LTMA * (1 − B) < STMA < LTMA * (1 + B). The reported t-statistics are for the two-tailed test that the mean net return differs from zero, and are computed from Newey and West (1987) standard errors. The number of markets with positive return and associated binomial probability excludes GI (the GS Commodity Index).

* Indicate statistical significance at the 5% level.

** Indicate statistical significance at the 1% level.

Table 4
Net returns by commodity to channel strategies.

Commodity	3-month channel		4-month channel		5-month channel		6-month channel		9-month channel		12-month channel	
	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat
BO	0.4659%	1.9049	0.4594%	1.9055	0.4007%	1.7257	0.3774%	1.6621	0.3653%	1.6872	0.3655%	1.7250
C	0.4241%	2.2303*	0.3169%	1.7605	0.2636%	1.5474	0.2306%	1.4340	0.2075%	1.3700	0.1340%	0.9859
CC	0.0126%	0.0432	0.0147%	0.0517	0.1326%	0.4894	0.1453%	0.5425	0.1772%	0.7262	0.2397%	1.0655
CL	0.3905%	1.7364	0.3722%	2.2621*	0.4198%	2.7392**	0.3589%	2.3719*	0.2590%	1.7732	0.3077%	2.1319*
CT	0.4015%	1.8533	0.3444%	1.7538	0.3826%	2.0188*	0.3767%	2.0203*	0.2678%	1.5578	0.2564%	1.5804
FC	0.2285%	1.5599	0.1528%	1.0858	0.1959%	1.4788	0.2149%	1.6807	0.1862%	1.6874	0.1967%	1.8558
GC	0.3184%	1.9217	0.3393%	2.2076*	0.3895%	2.6803**	0.3776%	2.6831**	0.2904%	2.2147*	0.2830%	2.2046*
HG	0.3583%	1.4609	0.3023%	1.2640	0.4200%	1.8766	0.4026%	1.7984	0.3548%	1.7523	0.3055%	1.5980
HO	0.2253%	1.0935	0.2017%	1.2661	0.1897%	1.1726	0.0461%	0.2827	0.0613%	0.4456	0.1076%	0.8229
HU	0.0751%	0.3172	0.1230%	0.5731	0.1357%	0.6467	0.1336%	0.6527	0.0563%	0.2908	0.0413%	0.2204
JO	0.0896%	0.3580	0.1127%	0.4656	0.0695%	0.2889	0.0707%	0.2972	−0.0084%	−0.0370	0.0563%	0.2778
KC	0.0586%	0.1761	−0.0085%	−0.0257	0.0279%	0.0888	−0.0416%	−0.1332	0.2158%	0.7608	0.2658%	0.9575
KW	0.3477%	1.7942	0.4023%	2.0624*	0.3699%	1.9136	0.3945%	2.0971*	0.2027%	1.1196	0.2669%	1.5849
LB	0.4246%	1.7192	0.2992%	1.2455	0.2870%	1.2208	0.3462%	1.5872	0.2574%	1.2047	0.3500%	1.6824
LC	0.2884%	1.8583	0.2421%	1.5953	0.2213%	1.5275	0.2210%	1.5915	0.2115%	1.6610	0.2169%	1.7469
LH	0.0642%	0.2602	0.0385%	0.1683	0.0688%	0.3166	0.0909%	0.4356	0.1055%	0.5263	0.1211%	0.6246
NG	0.4514%	1.7393	0.3070%	1.2258	0.2839%	1.2390	0.2267%	1.0646	0.0466%	0.2598	0.1623%	1.0263
O	0.0476%	0.1918	−0.0469%	−0.1936	−0.0405%	−0.1774	−0.0134%	−0.0647	0.1165%	0.6049	0.0580%	0.3108
PA	1.1497%	4.4801**	1.0204%	4.2108**	0.9350%	4.0975**	0.8664%	3.8150**	0.8057%	3.5989**	0.7337%	3.3135**
PB	0.1548%	0.4739	0.0481%	0.1531	0.0023%	0.0080	−0.0216%	−0.0768	−0.0798%	−0.3095	−0.0619%	−0.2854
PL	0.2331%	0.8661	0.1914%	0.7138	0.1608%	0.6092	0.1512%	0.6072	0.2523%	1.0389	0.1865%	0.7784
RR	0.3930%	1.7942	0.4441%	2.1085	0.3759%	1.9144	0.3866%	1.9771*	0.2777%	1.4984	0.2372%	1.8626
S	0.1481%	0.6821	0.2076%	1.0164	0.2596%	1.3138	0.1929%	0.9914	0.1537%	0.7882	0.1419%	0.7790
SB	1.1959%	2.8284**	1.3608%	3.4720**	1.1858%	3.0473**	1.0895%	2.7527**	0.9625%	2.6123**	0.8061%	2.2572
SE	0.0789%	1.4768	0.0508%	1.0064	0.0517%	1.0405	0.0561%	1.1367	0.0404%	0.9166	0.0134%	0.3390
SI	0.1562%	0.5232	0.2217%	0.7593	0.1902%	0.6350	0.1663%	0.5728	0.2304%	0.8650	0.2935%	1.1178
SM	0.4606%	1.7396	0.5452%	2.1588	0.4991%	2.0487	0.3874%	1.6143	0.4254%	1.9184	0.3221%	1.4807
W	0.1616%	0.7457	0.2096%	1.0250	0.1700%	0.8467	0.1591%	0.8311	0.1530%	0.8345	0.1884%	1.1210
GI	0.1461%	1.5922	0.1317%	1.4181	0.1420%	1.5477	0.1327%	1.4837	0.1345%	1.5141	0.1451%	1.6733
Number of markets with positive mean return	28		26		27		25		26		27	
Binomial prob.	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000	

Notes: All returns are unlevered monthly returns net of assumed transactions costs. The holding period for each strategy is assumed to be 1 month, i.e. position indicators are updated at the end of each calendar month and a new position in each commodity is taken for the following month. A long position is taken if the latest end-of-month unit value exceeds the maximum of the end-of-month unit values over the previous *L* months, and a short position if the latest unit value is less than the minimum of the end-of month unit values over the previous *L* months. If the latest unit value is between the minimum and maximum observed over the previous *L* months, no position is taken in the commodity. The reported *t*-statistics are for the two-tailed test that the mean net return differs from zero, and are computed using Newey and West (1987) standard errors. The number of markets with positive return and associated binomial probability excludes GI (the GS commodity index).

* Indicate statistical significance at the 5% level.

** Indicate statistical significance at the 1% level.

terization results in positive net mean returns for 22 commodities, and binomial tests indicate that the probability of observing this many commodities with positive returns is very small. For the channel strategies, the worst specification (i.e. the 6-month channel) results in 25 commodities with positive mean returns, with even lower probability values for the binomial sign test. As in the case of the momentum results, we are able to report statistically significant positive mean returns for only a minority of commodities in each specification of the DMAC and channel strategies. Nevertheless, the strong tendency of these strategies to produce positive returns across our broad array of commodities is impressive. This point should be kept in mind in assessing the likely robustness of the portfolio results reported in the next section.

4. Basic tests and results for portfolios of commodities

We conduct aggregate tests of the magnitude and significance of the returns arising from the trend-following trading strategies described in the previous section in two ways. First, we examine if the trend-following trading rules yield significant positive mean net returns when they are applied to the GS commodity index futures. These results, reported toward the bottom of Tables 3 and 4 (ticker symbol GI), are ambiguous. On the one hand, all of the 12 trading rules examined yield positive mean returns in the GI futures; however, only four DMAC specifications (and none of the channel rules) achieve statistical significance. This approach, however, is conceptually problematic, because momentum is generally assumed to be a security-specific rather than a market-wide effect. While some commodity pairs have highly correlated returns, many others do not; thus, it is very possible that returns for individual commodities may be predictable even if returns for an aggregate index are not.

We conduct more sophisticated tests of whether investors can earn potentially abnormal returns via simultaneous implementation of trend-following trading rules in many markets by constructing equally-weighted return portfolios for each specification of each rule across commodity futures markets. This approach is very different from the tests conducted on the GS index because trading signals are generated, and positions are taken, at the level of individual commodities, rather than on an aggregate basis. Thus, the pooled results reported below would be obtained by simultaneously taking long positions in some commodities, and short or neutral positions in others, based on their past price histories. We note, however, that the equal weighting of returns across commodities in our pooled results implies that, each month, an equal dollar amount would need to be invested in each market in which a position (long or short) is taken in order to earn the pooled returns that we report in Tables 5 and 6 below. Because notional contract size (in dollar terms) varies widely, the equal-weighting strategy requires large variations in the number of contracts traded in each market; thus (due to discrete contract sizes) the returns we report, while strongly indicative, may not be exactly replicable.

Portfolio returns arising from pooling returns across all 28 of the individual commodities examined in this study are reported in Table 5. We report returns for the entire July 1959–December 2007 sample period, as well as for four approximately equal length subperiods: 1959–1971, 1972–1983, 1984–1995 and 1996–2007. In Panel A of Table 5, we report pooled returns for the six momentum rules in Table 2; pooled returns for the DMAC strategies are reported in Panel B of Table 5, and pooled returns for channel strategies in Panel C. The Sharpe ratios reported in Table 5 are defined in annual terms; thus, Sharpe ratio = $(12\mu/\sqrt{12}\sigma)$, where μ is the mean monthly logarithmic excess return and σ is the standard deviation of monthly log excess returns.

All of the momentum strategies result in pooled mean returns over the entire sample that are significantly positive at the 1% level using Newey–West standard errors (with five lags used for the full sample period and four lags for the subperiods) and a two-tailed test. The pooled unlevered mean net excess returns range from about 0.33% to 0.49% per month, and the Sharpe ratios between 0.42 and 0.64.[11] For the first three subperiods, mean returns to the momentum strategies appear to be roughly comparable to those reported for the full sample period; the *t*-statistics tend to be somewhat lower (due to fewer observations), albeit still significant at the 5% level or better for at least three formation periods in each of the first three subperiods. For the final 1996–2007 subperiod, there are strong indications that the mean returns of the momentum strategies are lower. Only four of the six formation periods result in positive returns, and no formation period results in a statistically significant mean return at the 5% level.

Looking at the DMAC results in Panel B, and the channel results in Panel C, it appears that these strategies work even better, and more consistently, than the momentum strategies. The mean returns for both strategies are significantly positive not only for the entire sample period, but for two subperiods (1972–1983 and 1984–1995), regardless of parameterization. For the 1959–1971 period, the channel strategies produce uniformly positive and significant returns, while 3 of the 6 DMAC parameterizations result in significant positive returns. Even for the 1996–2007 period, where the momentum strategies did not perform well, 3 of 6 channel strategies, and 5 of 6 DMAC parameterizations, produce significantly positive mean net returns. The Sharpe ratios resulting from the DMAC and channel strategies are also higher, on average, than those of the momentum rules, both for the entire sample and for most subperiods.

We next examine trading rule returns pooled only across those 20 markets that exhibit relatively high volume, where the strategies are more likely to be implementable on a large scale. This subset excludes the eight commodities with the lowest overall trading volumes reported in Table 1: cocoa, orange juice, lumber, oats, palladium, platinum, rough rice, and domestic sugar. With these exclusions, each remaining commodity futures market has average daily trading volume of at least 88.76 million dollars over the entire sample period. The pooled results for this high-volume subset are reported in Table 6. Comparing these results for the entire sample period for all strategies (momentum, DMAC and channel) with those reported earlier in Table 5, mean returns are generally similar, but Sharpe ratios are slightly lower if trading is restricted to the high-volume markets; nevertheless, all of the strategies continue to exhibit statistically significantly positive pooled returns. When we examine differences by subperiod, we similarly observe that for most strategies and subperiods, there is not a large difference in mean return depending on whether the results are aggregated

[11] The mean returns to momentum strategies in our study appear substantially lower than those reported in either Shen et al. (2007) and Miffre and Rallis (2007) for two reasons. First, our study focuses on returns net of transactions costs, while the prior studies report gross returns. A more subtle, yet potentially more significant difference is the weighting scheme employed. The prior studies are primarily concerned with replicating stock market momentum studies in commodity futures; consequently, they assume a zero investment weighting scheme whereby past "loser" commodities are sold short and past "winner" commodities purchased with the proceeds. Because, in futures markets, one cannot sell short commodities in such a way as to obtain use of short sale proceeds, and thus separate margin accounts must be posted for all positions, this weighting scheme implicitly assumes that 50% of the initial notional contract value is held in each margin account. In contrast, because we use unlevered returns and equal weighting across commodities, we assume 100% margins in all markets in which positions are taken each month. Consequently, our reported mean returns are much lower, but the standard deviations of our returns are also commensurately lower. The *t*-statistics and Sharpe ratios among the studies are thus more directly comparable; these should differ only by the assumed level of transactions costs in our study.

Table 5
Pooled returns to trend-following strategies, implemented with all commodities.

	Entire sample period			1959–1971			1972–1983			1984–1995			1996–2007		
	Mean Return (%)	*t*-stat	Sharpe ratio	Mean Return (%)	*t*-stat	Sharpe ratio	Mean Return (%)	*t*-stat	Sharpe ratio	Mean Return (%)	*t*-stat	Sharpe ratio	Mean Return (%)	*t*-stat	Sharpe ratio
Panel A: Momentum strategies															
1-month FP	0.3311	2.9944**	0.4335	0.2279	1.2387	0.3543	0.5321	2.0691*	0.5963	0.7582	4.2177**	1.0174	−0.1907	−0.9006	−0.2586
2-month FP	0.4946	4.5792**	0.6448	0.5637	2.7124**	0.8040	0.7791	3.7926**	0.8783	0.6903	3.6833**	0.9400	−0.0563	−0.2603	−0.0785
3-month FP	0.4857	4.3312**	0.6444	0.5161	2.5046*	0.7096	0.5495	2.0358*	0.5847	0.4849	2.5718*	0.7454	0.3921	1.9131	0.5853
6-month FP	0.3425	3.3808**	0.4572	0.5116	2.6450**	0.7174	0.3998	1.5363	0.4310	0.1990	1.1901	0.3264	0.2607	1.4950	0.3652
9-month FP	0.3298	2.8862**	0.4208	0.3322	1.4241	0.4399	0.4092	1.4958	0.4539	0.3634	2.0253*	0.4833	0.2145	1.0743	0.2980
12-month FP	0.4001	3.6264**	0.5279	0.3990	2.0931*	0.5463	0.4170	1.6072	0.4788	0.3451	1.9143	0.5377	0.4392	1.9211	0.5655
Panel B: Dual moving average crossover strategies															
ST = 1, LT = 6, *B* = .025	0.7760	6.0393**	0.8589	0.8541	2.8958**	0.9304	1.1769	4.2670**	1.0066	0.7218	4.1206**	1.0513	0.3511	1.6106	0.4645
ST = 1, LT = 12, *B* = .05	0.7405	5.7572**	0.7695	0.5468	2.1136*	0.5118	1.1151	3.5541*	0.9401	0.7012	3.6288**	0.9347	0.5908	2.6376**	0.7600
ST = 2, LT = 6, *B* = .025	0.6175	4.3954**	0.6497	0.5043	1.4364	0.4897	0.9250	3.0820**	0.7573	0.5035	2.5423*	0.7162	0.5370	2.5266*	0.7083
ST = 2, LT = 12, *B* = .05	0.6196	4.4591**	0.6026	0.5480	1.7309	0.4521	0.9012	2.7079**	0.7213	0.4531	2.2562*	0.5876	0.5734	2.7039**	0.7200
ST = 1, LT = 6, *B* = 0	0.5255	5.1938**	0.7158	0.3513	1.8150	0.5540	0.9724	3.8004**	0.9289	0.4572	3.3137*	0.8254	0.3211	2.0916*	0.5515
ST = 1, LT = 12, *B* = 0	0.5329	5.3803**	0.7352	0.3500	2.1625*	0.5801	0.9147	3.3880**	0.8913	0.4294	2.9487**	0.7462	0.4300	2.7285**	0.7363
Panel C: Channel strategies															
3-month channel	0.7135	5.8077**	0.7950	0.5708	2.3901*	0.7008	1.2379	4.2613**	1.0602	0.7937	5.0345**	1.0906	0.2534	1.1227	0.3165
4-month channel	0.7699	6.1014**	0.8066	0.5510	2.0678*	0.6102	1.2665	4.3519**	1.0362	0.9098	5.4591**	1.1922	0.3541	1.5380	0.4139
5-month channel	0.8822	6.4342**	0.8506	0.7795	2.4792*	0.7303	1.2800	4.1455**	0.9852	0.9579	5.4406**	1.1972	0.5114	1.9897*	0.5610
6-month channel	0.8688	5.9919**	0.8009	0.7126	2.1202*	0.6563	1.3305	4.0693**	0.9588	0.9560	5.1558**	1.1488	0.4742	1.7958	0.4994
9-month channel	0.9945	6.1171**	0.8161	0.8142	2.2196*	0.6629	1.5330	4.0094**	0.9817	1.0412	4.8898**	1.1416	0.5834	2.0620*	0.5436
12-month channel	1.1424	6.2317**	0.8565	1.0551	2.3513*	0.7410	1.4846	3.5332*	0.8856	1.2185	5.3256*	1.2606	0.8065	2.5824*	0.6880

Notes: All returns are unlevered monthly excess returns net of assumed transactions costs. The momentum returns reported in Panel A for each formation period (FP) arise from an equal weighting, in each month, of returns in individual commodity futures markets in which positions are taken, using the corresponding FP in Table 2. Similarly, the DMAC returns in Panel B, and the channel returns in Panel C, arise from equally-weighted portfolios across commodities of various parameterizations of these trading rules as reported in Tables 3 and 4, respectively. The *t*-statistics are for the two-tailed test that the mean net return differs from zero, and are computed from Newey and West (1987) standard errors.

* Indicate statistical significance at the 5% level.

** Indicate statistical significance at the 1% level.

Table 6
Pooled returns to trend-following strategies, implemented with high-volume commodities only.

	Entire sample period			1959–1971			1972–1983			1984–1995			1996–2007		
	Mean Return (%)	*t*-stat	Sharpe ratio	Mean Return (%)	*t*-stat	Sharpe ratio	Mean Return (%)	*t*-stat	Sharpe ratio	Mean Return (%)	*t*-stat	Sharpe ratio	Mean Return (%)	*t*-stat	Sharpe ratio
Panel A: Momentum strategies															
1-month FP	0.3412	2.8125**	0.3889	0.4885	2.5261*	0.6645	0.3719	1.2939	0.3604	0.7589	3.6698**	0.9327	−0.2585	−1.1201	−0.2914
2-month FP	0.5451	4.4296**	0.6089	0.8209	3.7814**	1.0534	0.8397	3.3503**	0.8056	0.6052	2.6573**	0.7278	−0.0910	−0.3787	−0.1022
3-month FP	0.5032	3.8779**	0.5596	0.7147	3.0075**	0.8536	0.7776	2.4412	0.7033	0.4044	1.8536	0.5046	0.1134	0.5007	0.1392
6-month FP	0.3325	2.7414**	0.3771	0.6886	3.0069**	0.9087	0.4486	1.5105	0.4205	0.1741	0.8394	0.2154	0.0213	0.1074	0.0249
9-month FP	0.2579	2.0405	0.2894	0.5177	2.2143*	0.6695	0.4374	1.4189	0.4146	0.2283	1.0774	0.2648	−0.1446	−0.6870	−0.1716
12-month FP	0.3627	3.1187**	0.4285	0.5231	2.6398**	0.7075	0.2267	0.7727	0.2188	0.3421	1.7410	0.4571	0.3666	1.6722	0.4421
Panel B: Dual moving average crossover strategies															
ST = 1, LT = 6, *B* = .025	0.8182	5.5043**	0.7514	1.0786	3.2532**	0.9733	1.2653	3.7122**	0.8954	0.5843	3.1658**	0.7162	0.3465	1.4121	0.3799
ST = 1, LT = 12, *B* = .05	0.7978	5.4898**	0.7074	0.9378	3.2120**	0.7412	1.1648	3.2301**	0.8508	0.5919	2.8066**	0.6540	0.5027	2.0504*	0.5583
ST = 2, LT = 6, *B* = .025	0.6789	4.2585**	0.6046	0.8896	2.5087*	0.7564	1.0825	2.9206**	0.7501	0.3488	1.6171	0.4307	0.3961	1.5436	0.4141
ST = 2, LT = 12, *B* = .05	0.7056	4.5044**	0.5810	1.0595	2.9811**	0.7160	0.9097	2.3937*	0.6440	0.3871	1.7551	0.4172	0.4831	2.1063*	0.5116
ST = 1, LT = 6, *B* = 0	0.5206	4.3759**	0.5911	0.5080	2.4888*	0.6957	0.9920	3.0341**	0.7726	0.2985	2.0023*	0.4751	0.2837	1.6092	0.3957
ST = 1, LT = 12, *B* = 0	0.5375	4.7256**	0.6313	0.4919	2.6816**	0.7326	0.9618	2.9815**	0.7846	0.3418	2.2889*	0.5224	0.3527	2.0261*	0.4985
Panel C: Channel strategies															
3-month channel	0.7548	5.3325**	0.6971	0.7326	2.9504**	0.7981	1.3268	3.7810**	0.9356	0.6783	3.2669**	0.7466	0.2820	1.1095	0.2829
4-month channel	0.8471	5.8901**	0.7390	0.7566	2.7862**	0.7575	1.5067	4.3743**	1.0193	0.7692	4.0617**	0.8189	0.3563	1.2965	0.3304
5-month channel	0.9756	6.1346**	0.7880	0.9282	2.8471**	0.8130	1.6110	4.3215**	1.0201	0.8339	3.9956**	0.8348	0.5294	1.7774	0.4616
6-month channel	0.9616	5.7069**	0.7414	0.9556	2.7374**	0.8325	1.6348	4.0495**	0.9679	0.8264	3.8516**	0.7811	0.4296	1.4058	0.3623
9-month channel	1.0868	5.8589**	0.7607	1.1113	2.7534**	0.8559	1.7719	3.8966**	0.9935	0.8355	3.4699**	0.7107	0.6341	2.0405*	0.4612
12-month channel	1.2399	6.0838**	0.8022	1.2525	2.5293*	0.8268	1.6547	3.4223**	0.8681	1.1093	4.1697**	0.9047	0.9469	2.8639**	0.6453

Notes: All returns are unlevered monthly excess returns net of assumed transactions costs. The momentum returns reported in Panel A for each formation period (FP) arise from an equal weighting, in each month, of returns in individual commodity futures markets in which positions are taken, using the corresponding FP in Table 2. Similarly, the DMAC returns in Panel B, and the channel returns in Panel C, arise from equally-weighted portfolios across commodities of various parameterizations of these trading rules as reported in Tables 3 and 4, respectively. Only returns from the 20 markets with the highest average daily trading volumes during the entire sample period are pooled in this table. The reported t-statistics are for the two-tailed test that the mean net return differs from zero, and are computed from Newey and West (1987) standard errors.

* Indicate statistical significance at the 5% level.

** Indicate statistical significance at the 1% level.

across all 28 commodities or only the high-volume subset; once again, however, the Sharpe ratios (and the statistical significance of the mean returns) tends to be slightly lower for the high-volume subset, although this result is largely to be expected.[12] The most notable difference in results between Table 5 and 6 concerns the performance of momentum and DMAC strategies in the 1996–2007 period: the mean returns for all specifications of momentum and DMAC (but not channel) strategies are lower when implementation is restricted to high-volume commodities.

Although we do report some diminishment in the profitability of some of the technical trading rules in the 1996–2007 period, our results still strongly conflict with previous literature. In the conclusion of their exhaustive survey, Park and Irwin (2007) note that, in futures markets, technical trading strategies appeared to be profitable from the mid 1970s to the mid 1980s, but no study has yet comprehensively documented the profitability of these strategies after that period. Indeed, Park and Irwin (2005) test the 12 strategies employed by Lukac et al. (1988) – with some success – on pre-1984 data, and find that, over the 1985–2003 period these strategies almost uniformly generate losses. Similarly, Marshall et al. (2008), upon controlling for data-snooping bias, find no evidence of technical trading rule profitability in most markets during their 1984–2005 sample period. In contrast, in both Tables 5 and 6, we find that the profitability of DMAC and channel strategies persists during 1984–1995 and, to only a slightly lesser extent, during 1996–2007 as well. In the next section, we examine if our divergent findings are robust to different distributional assumptions and more pessimistic assumptions regarding transactions costs. We also address the likelihood that possible data-snooping bias could account for our results.

5. Robustness tests

The results in Table 1 clearly show that monthly returns to long positions in commodity futures are not normally distributed, yet the Newey and West (1987) *t*-tests used in the study assume normal distributions. To ascertain if these *t*-tests provide correct inferences we conduct a bootstrap procedure first developed by Brock et al. (1992), whereby we compare the mean excess returns, standard deviation of returns, and Sharpe ratios that arise from our DMAC and channel trading rules applied to the actual historical commodity returns series to those that are obtained when the same trading rules are applied to randomly generated monthly returns. We conduct 1000 replications, obtain empirical distributions of the mean returns, standard deviations of returns and Sharpe Ratios from these replications, compute probability values based on these empirical distributions, and compare these to the probability values implied by the Newey and West (1987) *t*-statistics reported earlier.

The key difficulty that arises in our study in the context of conducting the bootstrap procedure described above is that monthly returns across some of the commodities are highly correlated; these *contemporaneous* correlations across commodities are entirely consistent with an informationally efficient market and must be preserved in the bootstrapped series if correct inferences are to be drawn regarding the pooled returns that are the primary focus of this study.[13] However, preserving the contemporaneous correlations across commodities is not easy to do, because the various commodities begin trading at different points in time. To ensure that these correlations are preserved, we must break up each bootstrapping replication into numerous steps such that all groups of commodities are bootstrapped over periods where data is available for each commodity in the group. Specifically, we proceed as follows: between July 1959 and January 1961, we have data for eight commodities (ticker symbols BO, C-, CC, CT, O-, S-, SM and W-). For each month during this time period, in each replication, one month is drawn randomly, with replacement, from all available months (July 1959–December 2007), and returns for each of these eight commodities from the source month are assigned to the target month. Next, between February 1961 and November 1964, we have returns data for nine commodities (the eight listed previously plus SB); for each month during this time period one month is drawn randomly, with replacement, from all months between February 1961 and December 2007, and returns for each of the nine commodities from the source month are assigned to the target month. We keep proceeding in this manner every time data for a new commodity becomes available. The last iteration is the period May 1990–December 2007, by which time we have returns data for all 28 commodities, for which we draw randomly from all months between May 1990 and December 2007 and returns from each of the 28 commodities from the source month are assigned to the target month. We believe that the procedure just described eliminates any time-series dependence in commodity returns and maximizes the source data from which bootstrapped series are drawn, subject to the constraint that contemporaneous return correlations across commodities must be maintained.[14]

The results of the bootstrap procedure described above are reported in Tables 7 and 8. In Table 7, we report results for each of the six DMAC strategy parameterizations, and in Table 8 for each of the six channel strategies. In Panel A of each table, we report pooled results for all commodities over the entire sample period, and in Panel B pooled results for 20 high-volume commodities over the entire sample. In Panels C and D, respectively, we report pooled results for all commodities and for high-volume commodities for the 1996–2007 subperiod.[15] For each of the DMAC strategies in Table 7, and for each of the channel strategies in Table 8, we show actual results for the mean return (net of assumed transactions costs), the standard deviation of returns and the Sharpe ratio, bootstrap probability values based on 1000 replications, and Newey–West

[12] Our finding that standard deviations are generally higher (and Sharpe ratios and *t*-statistics lower) when results are aggregated across 20 rather than 28 commodities does not necessarily imply that the trading strategies we examine work less well in the high volume commodities. Even if the strategies worked equally well in low and high volume markets we would expect to observe these results because a portfolio that is constructed with fewer assets will, ceteris paribus, exhibit higher variance. This higher variance simultaneously results in a lower Sharperatio (by definition) and reduced power for statistical tests for the mean return.

[13] For the May 1990–December 2007 period, during which monthly returns for all 28 commodities are available, the highest correlations of monthly returns across commodities are observed within the petroleum complex (crude oil, heating oil, unleaded gasoline), within the soybean complex (soybeans, soybean meal, soybean oil), between #2 winter wheat and #2 soft red wheat, and between feeder and live cattle; not surprisingly, these correlations tend to be between 0.7 and 0.9. Somewhat lower, but still substantial positive correlations are observed between lean hogs and pork bellies, between natural gas and each of the petroleum contracts, among the various grains (corn, wheat, oats and soybeans), and among the precious metals (gold, platinum, silver). Economic linkages across commodity futures markets are examined more formally in Chng (2009).

[14] Since we have already controlled for microstructure effects by skipping at least one day between formation and holding periods and by ensuring no trading occurs when price limits are binding, the absence of time-series dependence in our measured commodity returns is, by definition, necessary for weak-form market efficiency to hold. This is the main reason we essentially apply only the simplest bootstrap methodology of Brock et al. (1992) whereby a random walk null model is used in the procedure. Additionally, the need to preserve contemporaneous correlations across commodities makes more complex null models, such as a GARCH-M model as used in Marshall et al. (2008), infeasible in our study.

[15] Because all 28 commodities trade between 1996 and 2007, conducting the bootstrap is considerably easier for this subperiod. We do so as follows: For each month during this time period beginning with January 1996, in each replication, one month is drawn randomly, with replacement, from all months between January 1996–December 2007, and returns for each of these 28 commodities from the source month are assigned to the target month.

Table 7
Bootstrap results for dual moving average crossover strategies.

	ST = 1, LT = 6, *B* = 0.025			ST = 1, LT = 12, *B* = 0.05			ST = 2, LT = 6, *B* = 0.025			ST = 2, LT = 12, *B* = 0.05			ST = 1, LT = 6, *B* = 0			ST = 1, LT = 12, *B* = 0		
	Mean Return	Standard deviation	Sharpe ratio	Mean Return	Standard deviation	Sharpe ratio	Mean Return	Standard deviation	Sharpe ratio	Mean Return	Standard deviation	Sharpe ratio	Mean Return	Standard deviation	Sharpe ratio	Mean Return	Standard deviation	Sharpe ratio
Panel A: Entire sample period, all commodities																		
Actual result	0.7760%	3.1295%	0.8589	0.7405%	3.3332%	0.7695	0.6175%	3.2924%	0.6497	0.6196%	3.5619%	0.6026	0.5255%	2.5431%	0.7158	0.5329%	2.5112%	0.7352
Bootstrap prob. value	1.000	0.965	1.000	1.000	0.982	1.000	1.000	0.970	1.000	1.000	0.997	1.000	1.000	0.866	1.000	1.000	0.802	1.000
Newey–West prob. value	1.000			1.000			1.000			1.000			1.000			1.000		
Panel B: Entire sample period, high-volume commodities																		
Actual result	0.8182%	3.7721%	0.7514	0.7978%	3.9069%	0.7074	0.6789%	3.8895%	0.6046	0.7056%	4.2066%	0.5810	0.5206%	3.0507%	0.5911	0.5375%	2.9496%	0.6313
Bootstrap prob. value	1.000	0.988	1.000	1.000	0.982	1.000	1.000	0.981	1.000	1.000	0.996	1.000	1.000	0.953	1.000	1.000	0.860	1.000
Newey–West prob. value	1.000			1.000			1.000			1.000			1.000			1.000		
Panel C: 1996–2007 period, all commodities																		
Actual result	0.3511%	2.6179%	0.4645	0.5908%	2.6926%	0.7600	0.5370%	2.6261%	0.7083	0.5734%	2.7584%	0.7200	0.3211%	2.0167%	0.5515	0.4300%	2.0231%	0.7363
Bootstrap prob. value	0.971	0.984	0.951	0.998	0.963	0.994	0.995	0.896	0.994	0.992	0.951	0.987	0.982	0.855	0.975	0.996	0.860	0.994
Newey–West prob. value	0.945			0.995			0.994			0.996			0.981			0.996		
Panel D: 1996–2007 period, high-volume commodities																		
Actual result	0.3465%	3.1594%	0.3799	0.5027%	3.1189%	0.5583	0.3961%	3.3135%	0.4141	0.4831%	3.2711%	0.5116	0.2837%	2.4842%	0.3957	0.3527%	2.4507%	0.4985
Bootstrap prob. value	0.936	0.963	0.914	0.973	0.792	0.970	0.950	0.962	0.927	0.960	0.853	0.952	0.949	0.840	0.935	0.968	0.803	0.959
Newey–West prob. value	0.920			0.979			0.938			0.982			0.945			0.978		

Notes: This table compares the mean returns, standard deviation of returns and Sharpe ratios obtained using each DMAC strategy on the actual data to empirical distributions of these values obtained using 1000 bootstraped replications. In each replication, monthly returns are drawn randomly (with replacement) across available time periods but together across commodities in order to preserve contemporaneous return relations. The Newey and West (1987) *t*-statistics for the mean returns reported in Table 5 and 6 are converted to prob. values and shown above for comparison. Assuming a two-tailed test, prob. values exceeding 0.975 and 0.995 indicate significant positive results at the 5% and 1% levels, respectively.

Table 8
Bootstrap results for channel strategies.

	3-month channel			4-month channel			5-month channel			6-month channel			9-month channel			12-month channel		
	Mean Return	Standard deviation	Sharpe ratio	Mean Return	Standard deviation	Sharpe ratio	Mean Return	Standard deviation	Sharpe ratio	Mean Return	Standard deviation	Sharpe ratio	Mean return	Standard deviation	Sharpe ratio	Mean Return	Standard deviation	Sharpe ratio
Panel A: Entire sample period, all commodities																		
Actual result	0.7135%	3.1089%	0.7950	0.7699%	3.3066%	0.8066	0.8822%	3.5925%	0.8506	0.8688%	3.7578%	0.8009	0.9945%	4.2214%	0.8161	1.1424%	4.6203%	0.8565
Bootstrap prob. value	1.000	0.601	1.000	1.000	0.548	1.000	1.000	0.756	1.000	1.000	0.769	1.000	1.000	0.850	1.000	1.000	0.924	1.000
Newey–West Prob. value	1.000			1.000			1.000			1.000			1.000			1.000		
Panel B: Entire sample period, high-volume commodities																		
Actual result	0.7548%	3.7511%	0.6971	0.8471%	3.9708%	0.7390	0.9756%	4.2888%	0.7880	0.9616%	4.4927%	0.7414	1.0868%	4.9490%	0.7607	1.2399%	5.3542%	0.8022
Bootstrap prob. value	1.000	0.790	1.000	1.000	0.733	1.000	1.000	0.852	1.000	1.000	0.857	1.000	1.000	0.877	1.000	1.000	0.921	1.000
Newey–West prob. value	1.000			1.000			1.000			1.000			1.000			1.000		
Panel C: 1996–2007 period, all commodities																		
Actual result	0.2534%	2.7735%	0.3165	0.3541%	2.9636%	0.4139	0.5114%	3.1578%	0.5610	0.4742%	3.2889%	0.4994	0.5834%	3.7178%	0.5436	0.8065%	4.0610%	0.6880
Bootstrap prob. Value	0.916	0.969	0.889	0.956	0.970	0.935	0.983	0.981	0.970	0.969	0.978	0.953	0.978	0.993	0.960	0.995	0.992	0.985
Newey–West prob. value	0.868			0.937			0.976			0.963			0.979			0.995		
Panel D: 1996–2007 period, high-volume commodities																		
Actual result	0.2820%	3.4534%	0.2829	0.3563%	3.7350%	0.3304	0.5294%	3.9730%	0.4616	0.4296%	4.1071%	0.3623	0.6341%	4.7623%	0.4612	0.9469%	5.0833%	0.6453
Bootstrap prob. value	0.883	0.970	0.856	0.907	0.981	0.875	0.964	0.983	0.938	0.914	0.981	0.884	0.958	0.991	0.915	0.994	0.978	0.981
Newey–West prob. value	0.865			0.902			0.961			0.919			0.978			0.998		

Notes: This table compares the mean returns, standard deviation of returns and Sharpe Ratios obtained using each channel strategy on the actual data to empirical distributions of these values obtained using 1000 bootstraped replications. In each replication, monthly returns are drawn randomly (with replacement) across available time periods but together across commodities in order to preserve contemporaneous return relations. The Newey and West (1987) *t*-statistics for the mean returns reported in Table 5 and 6 are converted to prob. values and shown above for comparison. Assuming a two-tailed test, prob. values exceeding 0.975 and 0.995 indicate significant positive results at the 5% and 1% levels, respectively.

probability values for comparison (the latter are calculated from the Newey–West t-statistics reported in Tables 5 and 6).

Quantitatively, the findings for the entire sample period in Panels A and B, in both tables, are very easy to summarize: in the case of every DMAC and channel strategy tested we find that the mean return and Sharpe ratio arising from applying these trading rules to the actual data is greater than the mean return or Sharpe ratio in any of the bootstrapped replications, implying that the bootstrap probability value is 1, and confirming that the inferences provided by the Newey–West t-statistics in Tables 5 and 6 for the mean returns are correct. We also observe that the return standard deviations arising from applying the trading rules to the actual data are generally higher than in the majority of the bootstrapped samples in which time-series dependence is removed: the probability values for the standard deviations range from 0.548 to 0.997, and are significant (i.e. greater than 0.975, assuming a 5% significance level and a two-tailed test) for two of the DMAC specifications when all commodities are pooled, and for four of the DMAC specifications when only high-volume commodities are pooled. These results suggest that the trading rules considered here do carry somewhat elevated levels of risk, taking positions in commodities at times of relatively high volatility; however, if the Sharpe ratio is a valid measure of the reward-to-risk ratio, then investors appear to be well-compensated for bearing this risk.

The results for the 1996–2007 period in Panels C and D of Tables 7 and 8 are decidedly more ambivalent as to whether the DMAC and channel strategies earned significant positive returns. Based on the bootstrap probability values, five of six specifications earned significant positive returns when results for all 28 commodities are pooled, and these specifications also have significant Sharpe ratios; however, none of these strategies achieve significant returns when only high-volume commodities are pooled. For the channel strategies, three specifications earn significantly positive returns when pooling across all commodities, but only one when pooling is restricted to high-volume commodity markets. These findings generally agree with those presented earlier, in Tables 5 and 6, that returns to most of the trend-following strategies appear to be lower between 1996 and 2007 than previously. Indeed, perhaps the most significant finding in Panels C and D of the tables is that, while the bootstrap and Newey–West probability values sometimes differ, the differences are relatively small and there is no indication that the Newey–West probability values are biased upward. In Panels C and D of Table 7, 7 of 12 Newey–West probability values for the mean return are actually lower than the corresponding bootstrap probability values, four are higher, and in one case the two probability values are equal. Similarly, in Panels C and D of Table 8, the Newey–West probability value for the mean is lower than the bootstrap probability value in 7 of 12 cases.

We next apply two methodologies for gauging the robustness of our results to possible data-snooping bias. The first is the reality-check approach developed by Sullivan et al. (1999), whereby we use our previously described bootstrapped replications to compare the maximum mean return, standard deviation and Sharpe ratio obtained using any of the 12 trading rules (six DMAC and six channel specifications), to the maximums in each of the 1000 bootstrapped replications, and construct empirical distributions for the maximum values. Our procedure closely follows (and is explained in more detail in) Marshall et al. (2008), except that the universe of trading rules examined in our study is much smaller.

The reality-check results, pooled across all commodities and across only high-volume commodities, and for the entire sample period (Panel A) and the 1996–2007 period (Panel B), are reported in Table 9. Irrespective of how we pool, for the entire sample period, these results show that the mean return and Sharpe ratio provided by the best trading rule (which happens to be the 12-month channel) is larger than the maximum across all 12 trading rules in any of 1000 bootstrapped replications. For the 1996–2007 period, our reality-check results are somewhat weaker, although, if we pool across all commodities, the bootstrap probability values in Panel B are significant for both the maximum mean return and the maximum Sharpe ratio. If we pool across high-volume commodities only, the probability value is still significant for the mean return, but not for the Sharpe ratio. Nevertheless, the overall results in Table 9 support the notion that our findings are unlikely to be due to data-snooping bias.

The weakness of the reality-check tests conducted above is the assumption that the entire universe of trading rules consists of the 12 that we test. While this may be technically true within the context of our study, it is difficult to defend this assumption given our familiarity with a large number of previous studies that have tested many more specifications of numerous trend-following rules. What if we, subconsciously, chose our DMAC and channel strategies because in previous studies these rules were found to perform best? Consequently, a more stringent indicator of whether our findings can be explained by data-snooping bias begins with the following question: Assuming that the best trading rule we report would in fact be the best not only among the 12 we test, but among a much larger universe of possible rules, how large could this universe be in order for the mean return accruing to our best-performing trading rule to still be significant at the 5% level after correcting for data-snooping bias?

Table 9
Reality check results.

	All commodities			High-volume commodities		
	Maximum mean return	Maximum Standard deviation	Maximum Sharpe ratio	Maximum mean return	Maximum Standard deviation	Maximum Sharpe ratio
Panel A: Entire sample period						
Actual result.	1.1424%	4.6203%	0.8589	1.2399%	5.3542%	0.8022
Bootstrap prob. value	1.000 **	0.920	1.000 **	1.000 **	0.919	1.000 **
Panel B: 1996–2007 period						
Actual result.	0.8065%	4.0610%	0.7600	0.9469%	5.0833%	0.6453
Bootstrap prob. value	0.997 **	0.994 *	0.977 *	0.993 *	0.987 *	0.948

Notes: The table compares the maximums of the mean returns, standard deviations of returns, and Sharpe ratios across 12 trading rules applied to the actual data (the six DMAC strategies in Table 7 and the six channel strategies in Table 8) to empirical distributions of these maximum values across the 12 trading rules applied to 1000 bootstrap replications. Our procedure follows Marshall et al. (2008), except that the universe of trading rules examined in our study is much smaller. As previously, in the bootstrap replications, monthly returns are drawn randomly across available time periods but together across commodities in order to preserve contemporaneous return relations across commodities.

* Indicate statistical significance at the 1% level assuming a two-tailed test.

** Indicate statistical significance at the 5% level assuming a two-tailed test.

When testing multiple hypotheses (or, in this case, trading rules) with the same data, a standard procedure is to apply the Bonferroni correction, whereby if one considers K trading rules, one would accept as significant any trading rule with a significance level less than α/K, where α is the desired experiment-wide significance level (usually 5%). However, the Bonferroni method assumes that the K hypotheses tested are independent (i.e. uncorrelated) with each other, which is definitely not the case in our study: the average correlation of the monthly returns between any two of our 12 trading rules is approximately 0.8. Thus, in the context of our study, the Bonferroni method is much too conservative. A Bonferroni-like methodology that performs better when hypotheses are highly correlated was developed by Tukey et al. (1985). As modified by Sankoh et al. (1997) based on Monte Carlo evidence, we calculate an adjusted critical α level (α_K) as follows:

$$\alpha_K = 1 - (1 - C\alpha)^{(1/\sqrt{K})}, \tag{3}$$

where C is a correction factor (we use 0.8 based on Table II in Sankoh et. al.), α is the desired overall type I error rate (we use 0.05), and K is the assumed number of trading rules in the universe. We actually solve Eq. (3) backward, i.e. we substitute for α_K the highest value based on the Newey–West t-statistics reported in Tables 5 and 6, and calculate a hypothetical universe size K.

For the entire sample period, pooling across all commodities, the highest t-statistic (6.4342) is provided by the 5-month channel rule. The equivalent α value is 0.000000000262; substituting this for α_K in Eq. (3) and solving for K yields a value for K that well exceeds 1 trillion, and roughly similar results are obtained for the entire sample period if pooling is restricted to high-volume commodities. Thus, we are confident in stating that data-mining is extremely unlikely to explain our full sample results. Once again, however, results for the 1996–2007 period are more equivocal. Pooling across all commodities, the best trading rule (1/12/0 DMAC) has a t-statistic of 2.7285 and equivalent α value of 0.007161; according to Eq. (3), a universe size of just 33 potential rules would now be sufficient to render the mean return for this rule insignificant. Similarly, a universe size of 72 potential rules would render the results for the best rule when pooling across high-volume commodities insignificant.

As previously motivated in our discussion of assumed transactions costs, we next examine whether our trading rule findings are robust to more pessimistic assumptions regarding these costs. Specifically, to mimic potential market-impact effects of trying to implement our strategies on a large scale, we now assume fixed transactions costs of \$20 per contract, and a bid-ask spread of three ticks in each market. We now therefore estimate monthly transactions cost as a percent of notional contract value as

$$TC_t = [20 + (3 \times \text{Tick Size} \times \text{CM})]/(\text{Price} \times \text{CM}), \tag{4}$$

where all variables are as defined previously. Mean net excess returns, Newey–West t-statistics and Sharpe ratios for DMAC and channel strategies implemented with all 28 commodities, using these more pessimistic assumptions, are reported in Table 10. For the full sample period, these results show that our one-trade-per month, intermediate horizon DMAC and channel strategies are robust to high transactions costs: every parameterization of both strategies continues to earn a significantly positive mean net excess return. For the four subperiods, findings are considerably less robust. Assuming high transactions costs, some of the DMAC strategies in Panel A earn negative returns in the 1959–1971 period, and while returns in the other subperiods are uniformly positive, they are uniformly significant only during 1972–1983. The channel rules (Panel B) fare better, as returns are uniformly positive for all parameterizations in all periods, and returns are uniformly significant during 1972–1983 and 1984–1995. We obtain positive mean

Table 10
Pooled returns to trend-following strategies, implemented with all commodities, assuming high transactions costs.

	Entire sample period			1959–1971			1972–1983			1984–1995			1996–2007		
	Mean Return (%)	t-stat	Sharpe ratio	Mean Return (%)	t-stat	Sharpe Ratio	Mean Return (%)	t-stat	Sharpe Ratio	Mean Return (%)	t-stat	Sharpe Ratio	Mean Return (%)	t-stat	Sharpe Ratio
Panel A: Dual moving average crossover strategies															
ST = 1, LT = 6, B = .025	0.5092%	3.9629**	0.5632	0.3280	1.1265	0.3592	0.9844	3.5744**	0.8418	0.5300	2.9955**	0.7698	0.1948	0.8941	0.2579
ST = 1, LT = 12, B = .05	0.4776	3.6602**	0.4955	0.0068	0.0262	0.0063	0.9224	2.9469**	0.7778	0.5268	2.7194**	0.7023	0.4349	1.9467	0.5594
ST = 2, LT = 6, B = .025	0.3538	2.4850*	0.3710	−0.0262	−0.0753	−0.0256	0.7335	2.4530*	0.6008	0.3270	1.6189	0.4625	0.3800	1.7816	0.5006
ST = 2, LT = 12, B = .05	0.3539	2.5472*	0.3445	−0.0029	−0.0094	−0.0024	0.7095	2.1390*	0.5680	0.2764	1.3756	0.3585	0.4177	1.9756*	0.5246
ST = 1, LT = 6, B = 0	0.2693	2.6184*	0.3655	−0.1318	−0.6906	−0.2085	0.7775	3.0535**	0.7431	0.2640	1.9204	0.4770	0.1673	1.0899	0.2873
ST = 1, LT = 12, B = 0	0.2844	2.8222**	0.3914	−0.1363	−0.8472	−0.2263	0.7194	2.6739**	0.7013	0.2609	1.7900	0.4533	0.2762	1.7673	0.4737
Panel B: Channel strategies															
3-month channel	0.4776	3.8221**	0.5298	0.0883	0.3710	0.1083	1.0454	3.6063**	0.8953	0.6943	4.4010**	0.9513	0.0983	0.4369	0.1229
4-month channel	0.5279	4.0873**	0.5496	0.0692	0.2601	0.0766	1.0745	3.6946**	0.8789	0.7832	4.5872**	1.0141	0.1986	0.8636	0.2322
5-month channel	0.6336	4.5553**	0.6085	0.2890	0.9242	0.2711	1.0894	3.5271**	0.8381	0.8060	4.4622**	0.9982	0.3562	1.3845	0.3906
6-month channel	0.6186	4.2118**	0.5683	0.2217	0.6620	0.2044	1.1391	3.4844**	0.8206	0.7941	4.2144**	0.9498	0.3189	1.2068	0.3357
9-month channel	0.7395	4.5256**	0.6063	0.3076	0.8461	0.2520	1.3429	3.5162**	0.8597	0.8637	4.0371**	0.9426	0.4298	1.5189	0.4002
12-month channel	0.8887	4.8499**	0.6664	0.5432	1.2240	0.3833	1.2953	3.0861**	0.7724	1.0437	4.5694**	1.0785	0.6536	2.0918*	0.5571

Notes: All returns are unlevered monthly excess returns net of assumed transactions costs, which are herein \$20 per contract and 3 ticks, and converted to percentage values using Eq. (4). The reported t-statistics are for the two-tailed test that the mean net return differs from zero, and are computed from Newey and West (1987) standard errors.

* Indicate statistical significance at the 5% level.

** Indicate statistical significance at the 1% level.

returns for all rules in the most recent 1996–2007 period, albeit these are generally not significant.

6. Conclusion

Previous studies document that momentum strategies earn significant abnormal returns in a wide variety of markets, including commodity futures markets. The consensus in the literature is that the performance of these strategies is due primarily to time-series dependence in realized returns rather then cross-sectional differences in expected returns across securities or markets. Logically, therefore, pure trend-following strategies should perform as well as, or better than momentum strategies; however, outside of currency markets (which may be distorted by central bank intervention activity) there is little evidence in the existing literature that suggests these strategies are actually profitable in futures markets, particularly over the past two decades.

Our study attempts to resolve this conundrum by examining trend-following trading strategies in commodity futures using the research design of a momentum study. That is, in contrast to previous studies of technical trading rules, we study the performance of these strategies over intermediate horizons by using monthly data over a 48-year period, and we examine their simultaneous implementation over a broad array of 28 commodity futures markets. When examined in this framework, we find that virtually all parameterizations of the dual moving average crossover and channel rules that we test yield hugely significant profits over our full sample period and most subperiods examined; indeed, these pure trend-following strategies generally produce higher mean returns and Sharpe ratios than momentum strategies. While it is true that pooling results across markets is generally necessary in order to find a robust degree of statistical significance, the profitability of these strategies is quite widespread: every parameterization of the DMAC and channel strategy results in positive mean net excess returns in at least 22 out of 28 markets over the full sample period. When we split our sample period into four subperiods and/or when we restrict implementation of the trend-following strategies to 20 markets with relatively high trading volume, we continue to find that the strategies perform well, although there is some indication that mean returns are lower, and less significant, in the 1996–2007 period, especially when pooling is restricted to high-volume markets.

We conduct bootstrap tests showing that despite the non-normal distributions of monthly commodity futures returns, the inferences gleaned from Newey and West (1987) *t*-tests appear to be valid, in the sense that the probability values yielded by comparing the level of mean returns achieved by the trading strategies on the actual data to those in the bootstrap replications are generally similar to those implied by the Newey–West *t*-statistics. Using two different approaches, we also show that our full sample pooled results are extremely unlikely to be explained by data-snooping, although our 1996–2007 results are less robust in this regard. Finally, we also show that considerably more pessimistic assumptions regarding transactions costs, to account for the hypothetical market impacts that may result if these strategies were implemented on a large scale, reduce (but generally do not eliminate) the profitability of the trend-following strategies.

We do not know the precise market-impact effects of trying to implement trend-following strategies on a very large scale, and we do not directly test if exposure to risk factors could explain the high returns to these strategies. Therefore, despite very strong results, we do not claim that our study definitively shows that commodity futures markets are weak-form inefficient. Our results do strongly suggest, however, that trend-following strategies perform much better at intermediate horizons than over the very short horizons that have been the focus of previous studies. In other words, it is the *research design* associated with momentum studies, not anything particularly magical or special about the momentum rule per se, that causes findings in the previous literature to diverge. When pure trend-following and momentum strategies are evaluated in the same framework, the pure trend-following strategies perform quite well indeed.

References

Asem, E., 2009. Dividends and price momentum. Journal of Banking and Finance 33, 486–494.

Brock, W.A., Lakonishok, J., LeBaron, B., 1992. Simple technical rules and the stochastic properties of stock returns. Journal of Finance 47, 1731–1764.

Chan, K., Hameed, A., Tong, W., 2000. Profitability of momentum strategies in the international equity markets. Journal of Financial and Quantitative Analysis 35, 153–172.

Chen, J., Hong, H., 2002. Discussion of momentum and autocorrelation in stock returns. Review of Financial Studies 15, 565–573.

Chng, M.T., 2009. Economic linkages across commodity futures: hedging and trading implications. Journal of Banking and Finance 33, 958–970.

Conrad, J., Kaul, G., 1998. An anatomy of trading strategies. Review of Financial Studies 11, 489–519.

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

Followill, R.A., Rodriguez, A.J., 1991. The estimation and determinants of bid-ask spreads in futures markets. Review of Futures Markets 10, 1–11.

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

Grundy, B.D., Martin, J.S., 2001. Understanding the nature of the risks and the source of the rewards to momentum investing. Review of Financial Studies 14, 29–78.

Harris, R.D.F., Yilmaz, F., 2009. A momentum trading strategy based on the low frequency component of the exchange rate. Journal of Banking and Finance 33, 1575–1585.

Jarque, C.M., Bera, A.K., 1987. A test for normality of observations and regression residuals. International Statistical Review 55, 163–172.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65–91.

Jegadeesh, N., Titman, S., 2001. Profitability of momentum strategies: an evaluation of alternative explanations. Journal of Finance 56, 699–721.

Jegadeesh, N., Titman, S., 2002. Cross-sectional and time-series determinants of momentum returns. Review of Financial Studies 15, 143–157.

Korajczyk, R.E., Sadka, R., 2004. Are momentum profits robust to trading costs? Journal of Finance 59, 1039–1082.

Lesmond, D.A., Schill, M.J., Zhou, C., 2004. The illusory nature of momentum profits. Journal of Financial Economics 71, 349–380.

LeBaron, B., 1999. Technical trading rule profitability and foreign exchange intervention. Journal of International Economics 49, 125–143.

Locke, P.R., Venkatesh, P.C., 1997. Futures market transaction costs. Journal of Futures Markets 17, 229–245.

Lukac, L.P., Brorsen, B.W., 1990. A comprehensive test of futures market disequilibrium. Financial Review 25, 593–622.

Lukac, L.P., Brorsen, B.W., Irwin, S.H., 1988. A test of futures market disequilibrium using twelve different technical trading systems. Applied Economics 20, 623–639.

Marshall, B.R., Cahan, R.H., Cahan, J.M., 2008. Can commodity futures be profitably traded with quantitative market timing strategies? Journal of Banking and Finance 32, 1810–1819.

Miffre, J., Rallis, G., 2007. Momentum strategies in commodity futures markets. Journal of Banking and Finance 31, 1863–1886.

Moskowitz, T.J., Grinblatt, M., 1999. Do industries explain momentum? Journal of Finance 54, 1249–1290.

Neely, C.J., 2002. The temporal pattern of trading rule returns and exchange rate intervention: intervention does not generate technical trading profits. Journal of International Economics 57, 211–232.

Newey, W.K., West, K.D., 1987. A simple positive definite, heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Park, C.H., Irwin, S.S., 2005. The profitability of technical trading rules in U.S. futures markets: A data snooping free test. Working Paper, University of Illinois at Urbana-Champaign.

Park, C.H., Irwin, S.S., 2007. What do we know about the profitability of technical analysis? Journal of Economic Surveys 21, 786–826.

Rouwenhorst, G., 1998. International momentum strategies. Journal of Finance 53, 267–284.

Sankoh, A.J., Huque, M.F., Dubey, S.D., 1997. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. Statistics in Medicine 16, 2529–2542.

Shen, Q., Szakmary, A.C., Sharma, S.C., 2007. An examination of momentum strategies in commodity futures markets. Journal of Futures Markets 27, 227–256.

Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance, and the bootstrap. Journal of Finance 54, 1647–1691.
Szakmary, A.C., Mathur, I., 1997. Central bank intervention and trading rule profits in foreign exchange markets. Journal of International Money and Finance 16, 513–535.
Taylor, S.J., 1994. Trading futures using a channel rule: a study of the predictive power of technical analysis with currency examples. Journal of Futures Markets 14, 215–235.
Tukey, J.W., Ciminera, J.L., Heyse, J.F., 1985. Testing the statistical certainty of a response to increasing doses of a drug. Biometrics 41, 295–301.