Investigating Statistical Arbitrage in Commodity Markets

Viviana Fanelli

University of Bari

MDEF, Urbino, September 20, 2014

Outline

Outline

Efficient Market Hypothesis: Fama 1970

Market Efficiency implies stock prices fully reflect all publicly available information instantaneously, thus no investment strategies can systematically earn abnormal returns.

$\Downarrow$

Stock market anomalies are violations of Efficient Market Hypothesis in which consistently abnormal returns can be earned by on some investment strategies that are constructed based on potential market inefficiencies.

$\Downarrow$

The Joint Hypothesis problem refers to that testing for market efficiency is problematic, or even impossible. Any attempts to test for market (in)efficiency must involve an equilibrium asset pricing models.

A bit of History

Statistical Arbitrage: long horizon trading opportunity that generates riskless profits.

- 1980: Pairs Trading: Nunzio Tartaglia.
- 1990: Statistical Arbitrage (StatArb): statistical-mathematical models are used to individuate pricing inefficiencies and to profit from mispricing.
- 2002: Losing interest in StatArb models. Bad performance and poor confidence.
- 2005: New interest in StatArb models.

Definition: Bondarenko (2003)

- For Bondarenko (2003) the definition of statistical arbitrage opportunity derives from the concept of pure arbitrage opportunity. A pure arbitrage opportunity is a zero cost trading strategy by which gains are received with no possibility of losses. Instead, a statistical arbitrage opportunity is a zero-cost trading strategy for which the expected payoff is positive and the conditional expected payoff in each state of the economy is nonnegative, meaning that the strategy payoff can be negative in some elementary states, as long as the average payoff in each final state is nonnegative.

Definition: Bondarenko (2003)

A zero-cost trading strategy with a payoff $Z_T = Z(I_T)$ is called a statistical arbitrage opportunity if

1. $E[Z_T|I_0] > 0$, and
2. $E[Z_T|I_0^{\xi_T}] \geq 0$ for all ξ_T,

where $I_t^{\xi_T} := (I_t; \xi_T) = (\xi_1, ..., \xi_t; \xi_T)$ denotes the augmented information set which in addition to the market information at time-t also includes the knowledge of the final state of the economy.

Definition: Jarrow et al. (2003, 2005)

- For Jarrow et al. (2003, 2005) a statistical arbitrage is a long horizon trading opportunity that generates riskless profits. It is a natural extension of the trading strategies utilized in the existing empirical literature on anomalies. Statistical arbitrage is defined without any reference to any equilibrium model, therefore, its existence is inconsistent with market equilibrium and, by inference, market efficiency. The notion of statistical arbitrage enables the rejection of market efficiency without invoking the joint hypothesis of an equilibrium model. The joint hypothesis of an equilibrium model is replaced with an assumed stochastic process for trading profit.

Definition: Jarrow et al. (2003, 2005)

A statistical arbitrage is a zero initial cost, self-financing trading strategy with cumulative discounted trading profits $v(n)$ and incremental trading profits $\Delta v(n)$ such that:

- $v(0) = 0$
- $\lim_{n\to\infty} E^P[v(n)] > 0$
- $\lim_{n\to\infty} P(v(n) < 0) = 0$ and
- $\lim_{n\to\infty} \frac{Var[\Delta v(n)]}{n} = 0$ if $P[v(n) < 0] > 0, \forall n < \infty$ $[\lim_{n\to\infty} Var[\Delta v(n) | \Delta v(n) < 0] = 0]$

P is the statistical probability measure.

Financial meaning

Statistical fair-price long-term relationship between assets

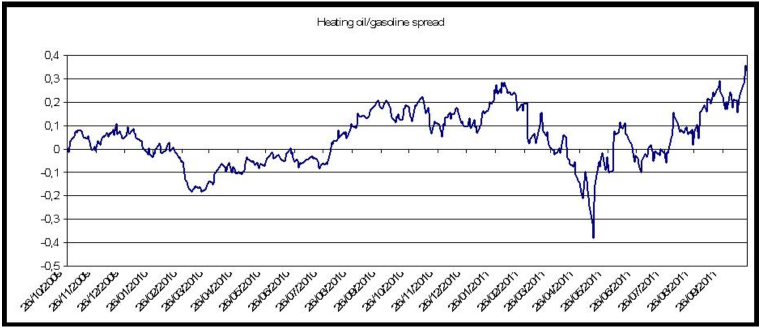
↻

Exploiting predictability in price dynamics

Intermarket Spreads

- Intermarket Spread: simultaneously buy and sell different, but connected commodities, that have a stable relationship
- There are intermarket spreads when commodities are substitute or whose prices are correlated
- Casual perturbations of supply and demand in the spot and futures markets give rise to intermarket spread opportunities when futures prices diverge from spot prices

Intermarket Spreads



Common Spreads

- Crack spread: the spread between crude oil price and refined products prices
- Frac spread: the spread between liquid gas and natural gas prices
- Spark spreads: the spread between electricity price and natural gas price

Assumptions

- Let $(\Omega, F, (F_t)_{t \geq 0}, P)$ be a filtered probability space over an infinite horizon $[0, \infty)$, satisfying the usual conditions. P is the statistical probability measure.
- Let the stochastic process $(M_t)_{t \geq 0}$ represent the discounted portfolio values on which a zero cost initial, self-financing trading strategy is built.

The portfolio $(M_t)_{t\geq 0}$

- Long-short Portfolio: it consists on a long(short) position on a target commodity and a short(long) position on a synthetic asset.
- It is a market neutral portfolio: the β of the portfolio is null, meaning that there is not correlation between the the return of the portfolio and the market return
- The assets are chosen by using the cointegration technique
- $(M_t)_{t\geq 0}$ follows a mean reverting dynamic
- $(M_t)_{t\geq 0}$ is called mispricing portfolio

The portfolio: the cointegration analysis

If two time series $(x_n)_{n \in \mathbf{N}}$ and $(y_n)_{n \in \mathbf{N}}$ are first-order integrated, but some cointegrating vector of coefficients exists to form a stationary linear combination of them,

$$y_n = \beta + \alpha x_n + \varepsilon_n, \; \varepsilon_n \sim N(0,1)$$

then they are cointegrated.

Cointegration: economic interpretation

- It can be seen as a long-period relationship: statistical fair-price relationship
- If x_t and y_t are linked by the same relationship, they can diverge in the short-term period, but the equilibrium will be restored
- There could be a statistical arbitrage

Description of the data

We consider a set of commodities, crude oils:

- The West Texas Intermediate (WTI), traded on the New York Mercantile Exchange (Nymex) was launched in March, 1983, and it is now the most liquid futures contract. The WTI is deliverable to Cushing, Oklahoma which is accessible to the spot market via pipeline.
- The Brent crude oil, which is traded on the Intercontinental Exchange (ICE) was launched in July, 1989.
- The Dubai crude oil quote by Platt's
- Time series: 25/10/2000-19/10/2009

Crude oil Mispricing Portfolio

We consider an infinite number of trading date t_i, with $i = 0, 1, ..., \infty$

The following statistical fair-price relationship holds for a generic t_i

$$E[T_{t_i}|\mathcal{F}_{t_i}] = Z_{t_i} = T_{t_i}, \qquad (1)$$

representing a long-term relationship among variables,
Z_{t_i} denotes the synthetic asset price at time t_i, the portfolio consisting of the commodities Brent and Dubai, T_{t_i} is the target price, that is The WTI crude oil.

Crude oil Mispricing Portfolio

The mispricing portfolio is obtained according to the following relationship

$$M_{t_i} = T_{t_i} - Z_{t_i}. \tag{2}$$

$$M_{t_i} = T_{t_i} - \beta_0 - \sum_{j=1}^{2} \beta_j C^j_{t_i}, \tag{3}$$

where $C^1_{t_i}$ is the price of the Brent oil, $C^2_{t_i}$ is the price of the Dubai oil.

Crude oil Mispricing Portfolio

Coefficients β_1, β_2 are estimated by using the cointegration regression:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
constant	1.617633	0.194177	8.330716	0.0000
BRENT	1.193785	0.0406112	9.39531	0.0000
DUBAI	-0.217020	0.042228	-5.139301	0.0000

R-squared	0.995153
Adjusted R-squared	0.995131
S.E. of regression	1.832739
Akaike info criterion	4.055987
Schwarz criterion	4.082885
F-statistic	47012.19
Prob(F-stat)	0.000000
Durbin-Watson statistic	0.256805

Table : Cointegration regression for crude oils.

Crude oil Mispricing Portfolio: Structural Break?

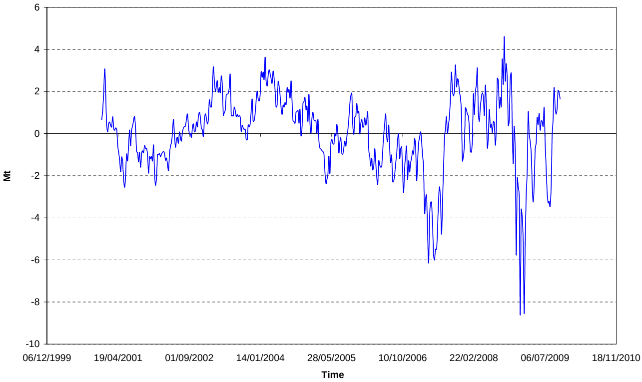
The QLR test of Stock and Watson (2003) provides a way to check whether the long run relationship between the WTI, Brent and Dubai crude oil is stable.

The QLR F-statistics tests the hyphothesis that the intercept and the coefficient in the equation (3) are constant against the alternative the break is in the central 70% of the sample. The critical value, F(3,455) = 61.2695, indicates that the null hypothesis that these coefficients are stable is rejected at the 1% significant level. This is evidence of structural change in the sample.

The breakpoint data is taken on Feb 14, 2005.

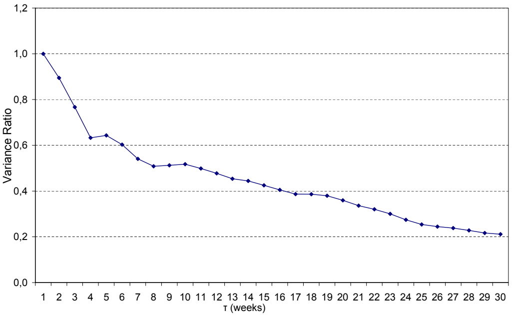
Finally, applying the Johansen (1991) test we verify that the cointegration relation holds also in presence of structural break.

Crude oil Mispricing Portfolio



Mean reverting investigation: Variance Ratio Function

The Augmented Dickey-Fuller test for mispricings is -3.75029 with a p-value of 0.04869



Mean Reverting Model

We use the Vasicek (1977) model under the statistical measure:

$$dM_t = \alpha(\theta - M_t)dt + \sigma dW_t, \quad M_0 = \overline{M} \tag{4}$$

where α, β, and σ are positive constants and W_t is a standard Wiener process.

Mean Reverting Model

With $0 \leq s \leq t$, we have:

$$M_t = \theta(1 - e^{-\alpha(t-s)}) + M_s e^{-\alpha(t-s)} + \sigma e^{-\alpha} \int_s^t e^{\alpha u} dW_u, \qquad (5)$$

M_t conditional to $\mathcal{F}_s$ is normally distributed with mean and variance given respectively by

$$E[M_t|\mathcal{F}_s] = \theta + (M_s - \theta)\, e^{-\alpha(t-s)} \qquad (6)$$

$$Var[M_t|\mathcal{F}_s] = \frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha(t-s)}\right) \qquad (7)$$

Mean Reverting Model: AR(1)

The discrete time version of the mean reverting model, on a time grid $0 = t_0, t_1, t_2, \ldots$ with (assume constant for simplicity) time step $\Delta t = t_i - t_{i-1}$, is:

$$M(t_i) = c + bM(t_{i-1}) + \delta\epsilon(t_i), \tag{8}$$

where the coefficients are:

$$c = \Theta(1 - e^{-\alpha \Delta t}),$$

$$b = e^{-\alpha \Delta t},$$

$$\delta = \sigma\sqrt{\Delta t},$$

and $\epsilon(t)$ is a Gaussian white noise $\epsilon \sim \mathcal{N}(0, 1)$.

Implicit Statistical Arbitrage Rules

- Implicit statistical arbitrage (ISA) trading strategies are based on trading rules that rely implicitly on the mean-reverting behaviour of the mispricing. Burgess (1999).
- If in the long time the misprisings reduce as prices change, an operator, who has previously opened a position, can realize profits provided that the impact of transaction costs, supported to employ the strategy, turns out less than the impact of the mispricing gain component.
- ISA trading rules define the sign and the magnitude of the mispricing portfolio components, that is the target commodity and the replication portfolio.

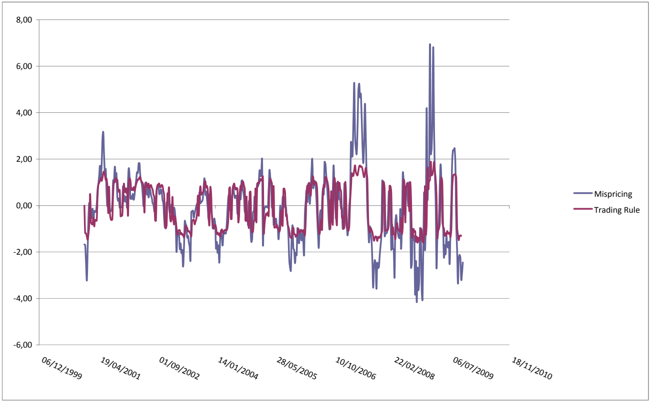
Implicit Statistical Arbitrage Rules

We indicate with $S(t_i, M_{t_i}, k)$ the basic trading rule at date t_i that depends on the sign and the level of the mispricing at the previous time and on the value of a sensitive parameter k according to the following formula

$$S(t_i, M_{t_i}, k) = -sign(M(t_{i-1}))|M(t_{i-1})|^k, \tag{9}$$

When $k = 0$ we have a step function meaning that the entire holding is always invested in the mispricing portfolio. If $k > 0$, the size of portfolio increases as the magnitude of the mispricing enlarges and, in particular, a $k > 1$ corresponds to more aggressive strategies.

Implicit Statistical Arbitrage Rules



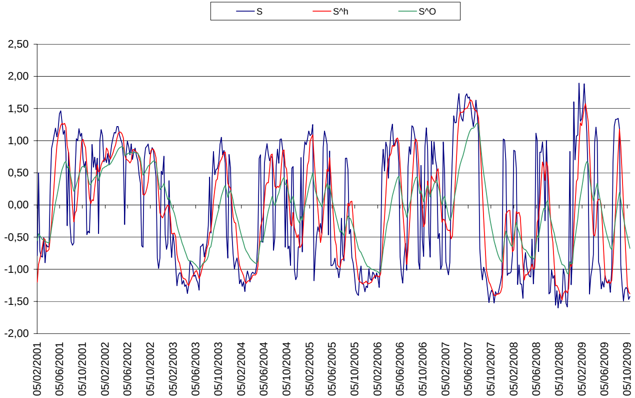
Implicit Statistical Arbitrage Rules

$$S^h(t_i, M_{t_i}, k, h) = S^h_{t_i} = \frac{1}{h}\sum_{j=1}^{h} S(t_{i-j}) \tag{10}$$

$$S^\Theta(t_i, M_{t_i}, k, \Theta) = S^\Theta_{t_i} = (1-\Theta)S_{t_i} + \Theta S^\Theta(t_{i-1}) \tag{11}$$

where $h \in \mathbf{N}$ is the moving average parameter and $\Theta \in \mathbf{R}_0^+$ is an exponential smoothing parameter.
By increasing the values of h and Θ, on one hand the number of transactions comes down so that the transaction costs decrease, on the other hand the accuracy of the smoothed trading signal diminishes.

Implicit Statistical Arbitrage Rules



Measuring the Performance

- The return gotten over a generic trading time period $[t_{i-1}, t_i]$:

$$R_{t_i} = S_{t_i} \frac{\Delta M_{t_i}}{T_{t_i} + Z_{t_i}} - c|\Delta S_{t_i}| \tag{12}$$

where $\Delta M_{t_i} = M_{t_i} - M(t_{i-1})$, $\Delta S_{t_i} = S_{t_i} - S(t_{i-1})$, c is the percentage transaction costs and $T_{t_i} + Z_{t_i}$ is the sum of the portfolio components.

- The total return or cumulative profit of a strategy:

$$cR_{t_i} = \sum_{j=1}^{i} R(t_j). \tag{13}$$

- The Sharpe Ratio: $Sh_{t_i} = \frac{\frac{1}{i}\sum_{s=1}^{i} R(t_s)}{\sqrt{\frac{1}{i}\sum_{s=1}^{i}\left[R(t_s) - \frac{1}{i}\sum_{s=1}^{i} R(t_s)\right]^2}}$

Verifying StatArb according to Jarrow et al. (2003)

Consider the discounted incremental cumulative profit of a strategy ΔcR_{t_i}, and assume they satisfy

$$\Delta cR_{t_i} = \mu t_i^{\gamma} + \sigma t_i^{\lambda} Z_i, \tag{14}$$

for $i = 1, ..., n$, where Z_i are i.i.d $\mathcal{N}(0,1)$ random variables with $Z_0 = 0$. Note that $\Delta cR(t_0) = 0$

Verifying StatArb according to Jarrow et al. (2003)

Theorem A trading strategy generates a statistical arbitrage with $1 - c$ percentage of confidence if the following conditions are satisfied:

1. $\mu > 0$
2. $\lambda < 0$ or $\gamma > \lambda$
3. $\gamma < \max\left\{\lambda - \frac{1}{2}; -1\right\}$.

with the sum of the p values for the individual tests forming an upper bound for the type I error c

Verifying StatArb according to Jarrow et al. (2003)

Parametro	Estimate	Standard Error	t Statistic	p-value
γ	−0,0108131	0,0781055	−0,1384	0,8899
λ	−0,746447	0,300529	−2,4838	0,0133

Mean of ΔcR_{t_i}	0,00131580
St. Dev. of ΔcR_{t_i}	0,0118019
Squared Residual Sum	0,0717678
Residual Standard Error ($\hat{\sigma}$)	0,0117820
R^2	0,00529009
Akaike criterion	−3135,0
Schwarz criterion	−3126,5

Out-of-Sample Analysis

Year	2001	2002	2003	2004
Total Return	7.76%	5.74%	4.76%	2.71%
Sharpe Ratio	1.67	1.42	1.14	0.98
Profitable weeks	45.45%	55.77%	42.31%	44.23%
Year	2005	2006	2007	
Total Return	2.09%	4.01%	13.89%	
Sharpe Ratio	0.90	0.86	0.83	
Profitable weeks	2.69%	50.00%	52.83%	

Table : In-sample performance

Year	2008	2009	2010
Total Return	4.16%	3.16%	4.12%
Sharpe Ratio	0.84	0.68	0.79
Profitable weeks	42.31%	59.52%	50.00%

Table : Out-of-sample performance

References

- Testing market efficiency using statistical arbitrage with applications to momentum and value strategies. S. Hogan, R. Jarrow, M. Teo, M. Warachka Journal of Financial economics, 2004 - Elsevier
- Statistical arbitrage and securities prices. Bondarenko O.- Review of Financial Studies, 2003