

Metodi statistici per le strategie di pair trading: un confronto

Minichini Carmine
0000892441

Overview

1. Cointegrazione

2. Metodi Statistici

3. Una strategia di Pair trading

3.1 Analisi di contegrazione

3.2 Definizione della strategia

3.3 Costruzione del portafoglio

4. Trading Performance: Un confronto tra metodi statistici

Overview

1. Cointegrazione

2. Metodi Statistici

3. Una strategia di Pair trading

3.1 Analisi di contegrazione

3.2 Definizione della strategia

3.3 Costruzione del portafoglio

4. Trading Performance: Un confronto tra metodi statistici

Processi Integrati

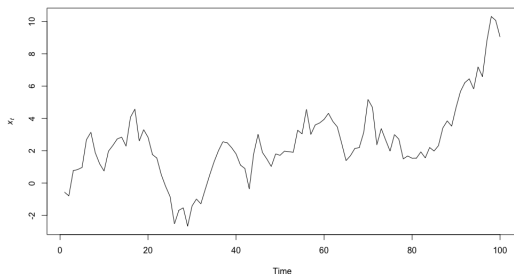


Figura: *Random Walk*

Una serie temporale x_t è definita integrata di ordine 1, $I(1)$, se non è stazionaria, ma è stazionaria nelle prime differenze $\Delta x_t = x_t - x_{t-1}$. L'esempio più semplice di processo $I(1)$ è il cosiddetto *random walk* o *passeggiata casuale*.

Cointegrazione

Il concetto di cointegrazione prende in esame due serie temporali: due serie temporali y_t e x_t , integrate di ordine $I(1)$, sono cointegrate se la loro combinazione lineare è un processo debolmente stazionario, $I(0)$. In altre parole il concetto di cointegrazione prende in esame la relazione tra due variabili, y_t e x_t ,

$$\epsilon_t = y_t - \beta x_t - \alpha \quad (1)$$

in cui la presenza del vettore di cointegrazione, $\hat{\beta}$, rende il processo ϵ_t , stazionario.

Engle-Granger two-step method

Generalmente l'approccio per testare se la combinazione di due serie storiche integrate produce un processo stazionario è quella di stimare dapprima i parametri α e β attraverso una regressione lineare, dopodichè testare la presenza di una radice unitaria all'interno del processo ϵ_t , per cui se $|\rho| < 1$ le due serie sono cointegrate.

$$y_t = \alpha_t + \beta_t x_t + \epsilon_t \quad (2)$$

$$\epsilon_t = \rho \epsilon_{t-1} + \eta_t \quad (3)$$

Per testare la presenza di radice unitaria nel processo ϵ_t è stata utilizzata la statistica test ADF (Augmented Dickey Fuller Test)

Overview

1. Cointegrazione

2. Metodi Statistici

3. Una strategia di Pair trading

3.1 Analisi di contegrazione

3.2 Definizione della strategia

3.3 Costruzione del portafoglio

4. Trading Performance: Un confronto tra metodi statistici

- ▶ **Approccio classico di cointegrazione**
- ▶ **Kalman Filter**
- ▶ **Rolling Regression**

Linear State Space Models

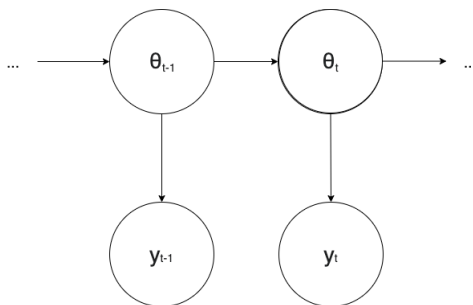


Figura: *Diagramma State Space Model*

L'idea metodologica alla base degli State Space Models è che lo sviluppo nel tempo del fenomeno in analisi, $y_1 \dots y_n$ è determinato da una serie di vettori non osservabili $\theta_1 \dots \theta_n$.

La relazione tra y_t e θ_t specifica il modello nella forma state space.

Gaussian Linear State Space Model

La relazione lineare tra due variabili cointegrate è descritta da:

$$y_t = \alpha + \beta x_t + \epsilon_t \quad (4)$$

In cui α e β sono invarianti nel tempo. La formulazione dell'equazione 4 come un modello state space Gaussiano ci permette di considerare invece l'evoluzione degli stati α e β nel tempo. Assumendo che questi si evolvano secondo un processo random walk avremo che ¹:

$$y_t = \alpha_t + \beta_t x_t + \epsilon_t \quad (5)$$

$$\alpha_t = \alpha_{t-1} + \eta_{1,t} \quad (6)$$

$$\beta_t = \beta_{t-1} + \eta_{2,t} \quad (7)$$

¹J. Durbin and S. J. Koopman, Time Series Analysis by State Space Methods, 2nd Ed. Oxford University Press, 2012

Gaussian Linear State Space Model

$$\theta_t = T_t \theta_{t-1} + \eta_t \quad (8)$$

$$y_t = Z_t \theta_t + \epsilon_t \quad (9)$$

- ▶ $\theta_t \triangleq \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix}$ è lo stato inosservato
- ▶ $T_t \triangleq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ è la matrice di transizione degli stati
- ▶ $\eta_t \sim \mathcal{N}(0, Q)$ è detto **state transition noise** ed è un errore gaussiano a media 0 e con matrice di covarianza $Q \triangleq \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$
- ▶ $Z_t \triangleq [1 \quad x_t]$ è la matrice di osservazione, dove x_t rappresenta, nel nostro caso, un vettore di prezzi.
- ▶ $\epsilon_t \sim \mathcal{N}(0, \sigma_e^2)$ è detto **observation noise**

Kalman Filter

Il termine Kalman Filter o Kalman Filtering fa riferimento ad una procedura ricorsiva che permette di conoscere la vera natura dello stato θ_t . La nozione chiave è che possiamo fare inferenza su θ_t attraverso una diretta applicazione del teorema di Bayes, in altre parole

$$Pr(Natura\ dello\ stato|Dati) \propto Pr(Dati|Natura\ dello\ stato) \quad (10)$$

che può essere scritto come

$$P(\theta_t|Y_t) \propto P(Y_t|\theta_t, Y_{t-1}) \cdot P(\theta_t|Y_{t-1}) \quad (11)$$

dove i due termini alla destra dell'equazione sono rispettivamente la verosimiglianza e la distribuzione a priori dello stato latente. Il termine sulla sinistra è la distribuzione a posteriori dello stato θ .

Overview

1. Cointegrazione

2. Metodi Statistici

3. Una strategia di Pair trading

3.1 Analisi di contegrazione

3.2 Definizione della strategia

3.3 Costruzione del portafoglio

4. Trading Performance: Un confronto tra metodi statistici

L'analisi empirica è stata svolta considerando i prezzi di chiusura giornalieri, aggiustati per i dividendi, dei costituenti dell'indice FTSE MiB nel periodo dal 1° gennaio 2015 al 31 dicembre 2019.²

Il campione ottenuto è stato poi diviso in:

- ▶ **Formation Period:** Dal 1° gennaio 2015 al 31 dicembre 2018 (4 anni)
- ▶ **Trading Period:** Dal 1° gennaio 2019 al 31 dicembre 2019 (1 anno)

²I prezzi di chiusura giornalieri, aggiustati per i dividendi, sono stati ottenuti tramite Yahoo Finance

Selezione delle coppie cointegrate

Nel *formation period*, per i 40 titoli dell'indice è stata analizzata la relazione di cointegrazione tramite l'Engle-Granger two-step method. Da un punto di vista pratico sono state calcolate quindi $\frac{n^2-n}{2}$ possibili combinazioni di titoli tra i 40 presi in oggetto, per un totale di 561 combinazioni.



Figura: *logaritmo dei prezzi*

Risultati della selezione

La Tabella mostra i risultati della selezione ottenuta mediante i passaggi precedentemente discussi.

| Pair | ρ | ADF Statistic | p-value | Settore |
|------------------|--------|------------------|---------|-------------|
| DIA.MI - AMP.MI | 0.97 | -3.66 | 0.026 | Sanitario |
| UCG.MI - BPE.MI | 0.94 | -3.73 | 0.022 | Bancario |
| BMED.MI - BGN.MI | 0.80 | -4.43 | 0.01 | Finanziario |
| BMED.MI - AZM.MI | 0.61 | -3.98 | 0.01 | Finanziario |

Definizione della strategia

Lo spread nel *trading period* è stato calcolato come:

$$z_t = \ln(stock_{1,t}) - \hat{\beta} \ln(stock_{2,t}) - \hat{\alpha} \quad (12)$$

La caratteristica di *mean reversion* del processo, ci permette di definire la strategia operativa che utilizzeremo.

Definizione della strategia

- ▶ Se $z_t > threshold$: significa che $stock_1$ è sopravvalutato rispetto a $stock_2$, per cui vendiamo 1 unità di $stock_1$ e compriamo $\hat{\beta}$ unità di $stock_2$
Chiudiamo la posizione quando $z_t \leq 0$
- ▶ Se $z_t < -threshold$: significa che $stock_{1,t}$ è sottovalutato rispetto a $stock_2$, per cui compriamo 1 unità di $stock_1$ e vendiamo $\hat{\beta}$ unità di $stock_2$
Chiudiamo la posizione quando $z_t \geq 0$

Costruzione del portafoglio

L'intera strategia di trading è equivalente ad avere un portafoglio con pesi³

$$w \triangleq \begin{bmatrix} 1 \\ -\hat{\beta} \end{bmatrix} \quad (13)$$

Sotto il vincolo che $\|\mathbf{w}\| = 1$ il portafoglio costruito per la strategia è :

$$w \triangleq \begin{bmatrix} 1/(1 + \hat{\beta}) \\ -\hat{\beta}/(1 + \hat{\beta}) \end{bmatrix} \quad (14)$$

mentre i rendimenti delle due posizioni sono stati poi calcolati come

$$w^T \Delta \ln(\text{stocks}) \quad (15)$$

Il rendimento totale del portafoglio è stato calcolato come il differenziale dei rendimenti tra le due posizioni.

³Yiyong Feng and Daniel P. Palomar (2016), "A Signal Processing Perspective on Financial Engineering"

Overview

1. Cointegrazione

2. Metodi Statistici

3. Una strategia di Pair trading

3.1 Analisi di contegrazione

3.2 Definizione della strategia

3.3 Costruzione del portafoglio

4. Trading Performance: Un confronto tra metodi statistici

Nel nostro lavoro di *Back Testing* della strategia è necessario però fare delle premesse, assumiamo cioè che valgano le seguenti condizioni

1. Nell'analisi della *trading performance* non esistono costi di transazioni, questo elimina l'onere di considerare fattori come il *bid-ask spread* , commissioni applicate dai broker e lotti minimi di negoziazione.
2. Nell'ipotesi di rottura del legame di cointegrazione non vi è un intervento nella strategia.

Misure di performance utilizzate

Sharpe Ratio:

$$SR = \frac{\mathbb{E}[r_t] - \mathbb{E}[r_{ft}]}{\sigma} \cdot \sqrt{252} \quad (16)$$

Max Drawdown:

$$MDD = \frac{(Max - Min)}{Max} \quad (17)$$

Compounded Returns:

$$prod(1 + R) - 1 \quad (18)$$

Value-at-Risk(VaR):

$$VaR_{\alpha}(X) = -\mu \cdot \Phi^{-1} \cdot \sigma \cdot \sqrt{252} \quad (19)$$

Expected Shortfall(ES):

$$ES_{\alpha}(X) = \frac{\phi(\Phi^{-1})}{\alpha} \cdot \sigma \cdot \sqrt{252} \quad (20)$$

La prima coppia considerata (in ordine di correlazione) è formata dalle multinazionali **DiaSorin S.p.A.** e **Amplifon S.p.A.**, entrambe operanti nel settore medico-sanitario.

| Processo | Metodologia | ADF Statistic | p-value | RMSE |
|----------|--------------------|------------------|---------|-------|
| S_1 | Cointegrazione | -3.55 | 0.041 | 0.037 |
| S_2 | Kalman Filter | -5.01 | 0.01 | 0.013 |
| S_3 | Rolling Regression | -3.70 | 0.020 | 0.032 |

| Metodologia | Sharpe Ratio | Max Drawdown | Returns | VaR | ES |
|--------------------|--------------|--------------|---------|------|------|
| Cointegrazione | 2.61 | 0.051 | 0.39 | 0.28 | 0.32 |
| Kalman Filter | 2.01 | 0.025 | 0.24 | 0.24 | 0.28 |
| Rolling Regression | 3.16 | 0.051 | 0.47 | 0.26 | 0.30 |

Osserviamo che i risultati del *VaR* e dell'*ES* ci suggeriscono che ad un livello di confidenza del 95% la metodologia che subisce minori perdite è il *Kalman Filter*.

La natura fortemente stazionaria dello spread calcolato con la metodologia del Kalman Filter, ci suggerisce che un portafoglio costruito con questo metodo è meno soggetto a variazioni estreme nella coda sinistra della distribuzione.

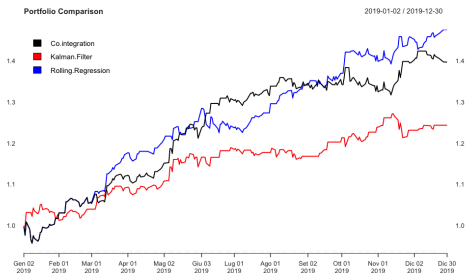


Figura: Equity line delle strategie

La seconda coppia considerata è formata dalle multinazionali **Banca Mediolanum S.p.A.** e **Azimut Holding S.p.A.**, entrambe operanti nel settore dei servizi finanziario-assicurativi.

| Processo | Metodologia | ADF Statistic | p-value | RMSE |
|----------|--------------------|------------------|---------|-------|
| S_1 | Cointegrazione | -0.69 | 0.96 | 0.11 |
| S_2 | Kalman Filter | -4.15 | 0.01 | 0.011 |
| S_3 | Rolling Regression | -1.99 | 0.57 | 0.04 |

Banca Mediolanum S.p.A. - Azimut Holding S.p.A.

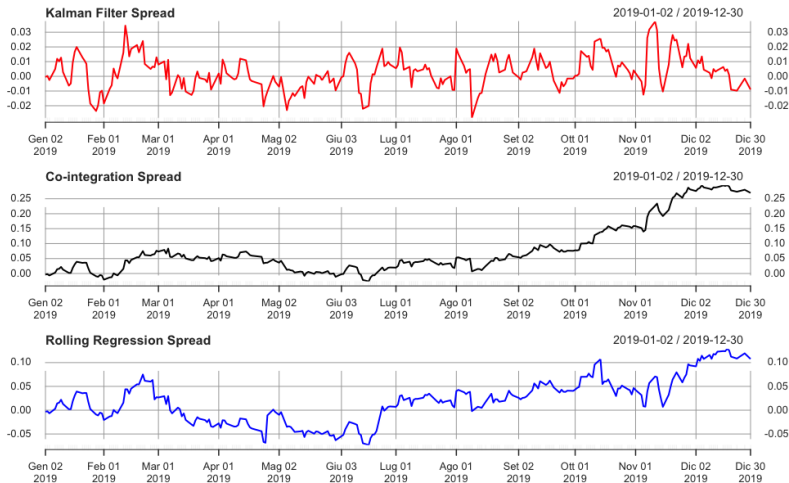


Figura: Spread per le tre metodologie

| Metodologia | Sharpe Ratio | Max Drawdown | Returns | VaR | ES |
|--------------------|--------------|--------------|---------|------|------|
| Cointegrazione | -0.79 | 0.18 | -0.08 | 0.26 | 0.30 |
| Kalman Filter | 0.72 | 0.04 | 0.06 | 0.06 | 0.23 |
| Rolling Regression | -1.00 | 0.13 | -0.10 | 0.24 | 0.28 |

Banca Mediolanum S.p.A. - Azimut Holding S.p.A.

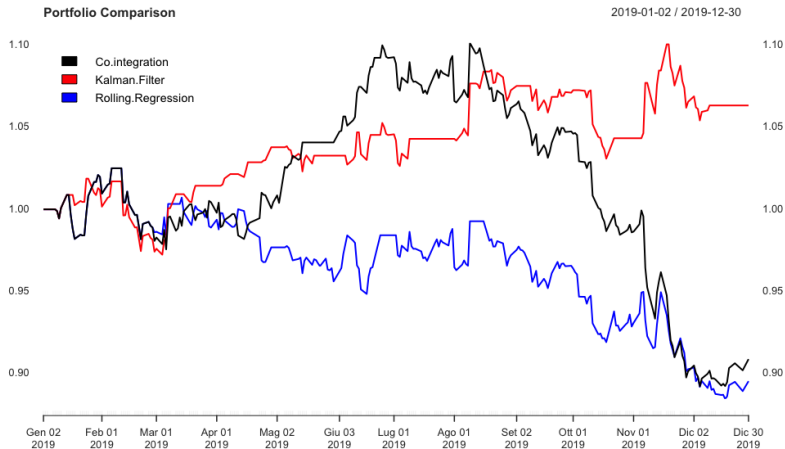


Figura: Equity line delle strategie