Non-Parametric Entropy

(I) Introduction to Measure and Test for Serial Dependence with Applications to Pairs Trading Strategies

Kuan-Lun WANG [1]

National Taiwan University
https://www.csie.ntu.edu.tw/~d06922002/

April 30, 2018

[0]This talk is based on [4] (midterm presentation for ITCT).
[1]E-mail: polyphonicared@gmail.com

Short Biodata

Kuan-Lun Wang is a doctoral student majoring in generalized pairs trading. The main goal of his research is to develop an algorithmic trading mechanism based on statistical arbitrage. His areas of expertise include automatic search procedures for model selection, multivariate co-integration approach, and structural change test.

Research Interests

Kuan-Lun Wang's research interests comprise time series models, simulation modeling, and portfolio choice. The central themes of his application are the study of multivariate pairs trading in real time, search for assets with a long-run equilibrium, and building of riskless portfolios. Much of his current work involves conducting structural change analysis and co-integration test of the finite order vector autoregressive process and estimating the probability of mean reversion. Such methods are important in a variety of applications, including economic indicators and hedging. One such application is index funds being tied to indexes with very low costs and risks.

Outline

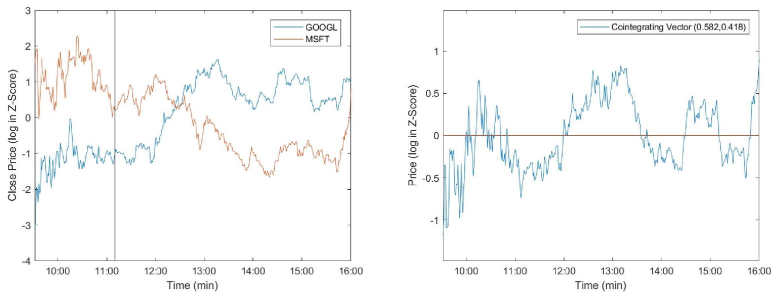
Should you have any questions, feel free to contact us.
(Don't believe me! Just check it by yourself!!)

Pairs Trading Strategies (1/5)

1. Pairs trading strategies (PTSs),
2. A popular short-term speculation strategy on Wall Street [2],
3. Two-step process:
    1. Find two assets with prices that move together, historically;
    2. Monitor the subsequent trading spread between the assets. Should the spread diverge, short the higher priced asset and long the lower priced asset until the spread converges again.

What are "move together" and "converges"?

Pairs Trading Strategies (2/5)



(a) Log Prices Process (b) Mean Reversion Process

Figure: An Example for PTSs

Pairs Trading Strategies (3/5)

1. Intent: equilibrium relationships [7, 1, 2, 5].
   - E.g., $\beta' y_t = \beta_1 y_{1t} + \cdots + \beta_K y_{Kt} + c = 0$.
2. Strategies: mean-reverting strategies.
   - Assumption: relatively large price moves will not continue,
   - Cost: less slippage than trend trading strategies.
3. Goal: statistical arbitrage.
   - Performance: outperform the market in high probability. [6]

Pairs Trading Strategies (4/5)

We consider a n-dimensional asset prices process:

$$y_t = f(y_{t-1}, y_{t-2}, \ldots) + u_t,\ \ y_t = \begin{pmatrix} y_t^{(1)} \\ \vdots \\ y_t^{(n)} \end{pmatrix}.$$

Because linear functions are relatively easy to deal with, they have been used so that, for example, vector autoregressive process:

$$\begin{aligned} y_t =& f(y_{t-1}, y_{t-2}, \ldots) && + u_t \\ & \nu + A_1 y_{t-1} + A_2 y_{t-2} + \cdots && + u_t. \\ & \nu + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} && + u_t. \end{aligned}$$

Of course, we can consider another nonlinear process.

Pairs Trading Strategies (5/5)

Assumption 1: Normality

Take $\{u_t\}$ to be normal process with time invariant covariance matrix Σ_u.

Assumption 2: Roots

The characteristic polynomial $A(z)$ satisfies the condition that if $|A(z)| = 0$, then either $|z| > 1$ or $z = 1$.

Assumption 3: VECM Representation

There are α and β such that $\alpha\beta' = (-(I - \sum_{i=1}^{p} A_i)', v')$.

Then we can estimate and test this vector equilibrium β, and $\beta' y_t$ is a strictly stationary and normal process. (This is my master thesis.)

Non-Parametric Entropy Measures (1/2)

Suppose $\{X_t\}$ is a strictly stationary time series with marginal density $g(\cdot)$ and pairwise joint density $f_j(\cdot)$ for $Z_{jt} \equiv (X_t, X_{t-j})'$, where $j \in \mathbb{N}$ is a given lag order.
We will show there is a isomorphic function from $\beta' y_t$ to X_t in final section.

Criterion: Kullback-Leibler Information Criterion (KLD)

To measure deviations of $f_j(\cdot)$ from $g(\cdot)g(\cdot)$, one can use the following criterion:

$$\mathcal{I}(j) \equiv \int \ln \left[\frac{f_j(x,y)}{g(x)g(y)}\right] f_j(x,y)\,\mathrm{d}\,x\,\mathrm{d}\,y,\ j \in \mathbb{N},$$

where the integral is take over the support of Z_{jt}.

Clearly, $\mathcal{I}(j) = 0$ if and only if X_t and X_{t-j} are independent.

Non-Parametric Entropy Measures (2/2)

Estimator: Kernel Estimator

Let X be a variable has the p.d.f. f, and $X_1, \ldots, X_n$ be the sample of X. For a bandwidth h_n ($\downarrow$) and a kernel k, kernel estimator can be defined as:
$$\hat{f}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h_n}\right).$$

One can estimate the KLD:

$$\hat{\mathcal{I}}_n(j) \equiv \frac{1}{n-j}\sum_{t \in S_n(j)} \ln\left[\frac{\hat{f}_{jt}(Z_{jt})}{\hat{g}_t(X_t)\hat{g}_{t-j}(X_{t-j})}\right],\ j = 1, \ldots, n-1,$$

where $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$ are kernel estimators for $f_j(\cdot)$ and $g(\cdot)$, and $S_n(j) \equiv \{t \in \mathbb{N} : j < t \le n, \hat{f}_{jt}(Z_{jt}) > 0, \hat{g}_t(X_t) > 0, \hat{g}_{t-j}(X_{t-j}) > 0\}$ [3].
However, no limiting distribution theory is available for $\hat{\mathcal{I}}(j)$ [3].

Heuristic on the Asymptotic Expansion of the Entropy Estimators (1/8)

To derive the limiting distribution of $\hat{\mathcal{I}}_n(j)$, we will Taylor-expand $\hat{\mathcal{I}}_n(j)$ up to a second order and show that the first two terms jointly determine the limit distribution of $\hat{\mathcal{I}}_n(j)$.

Assumption A.1

Take $\{X_t\}$ to be strictly stationary with X_t having support on $\mathbb{I} \equiv [0,1]$. Its marginal density $g : \mathbb{I} \to \mathbb{R}^+$ exists, is bounded away from 0, and is continuously twice differentiable on $\mathbb{I}$. Moreover, $|g^{(2)}(x_1) - g^{(1)}(x_2)| \leq C|x_1 - x_2|$ for any $x_1, x_2 \in \mathbb{I}$.

Assumption A.2

Assume $k : [-1,1] \to \mathbb{R}^+$ is a symmetric bounded probability density function.

Pairs Trading Strategies | Non-Parametric Entropy Measures | Testing the i.i.d. Hypothesis | References
Heuristic on the Asymptotic Expansion of the Entropy Estimators

Heuristic on the Asymptotic Expansion of the Entropy Estimators (2/8)

We write

$$
\begin{aligned}
\hat{\mathcal{I}}_n(j) =& \frac{1}{n_j} \sum_{t \in S_n(j)} \Bigg\{ \ln\left[\frac{f_j(Z_{jt})}{g(X_t)g(X_{t-j})}\right] + \ln\left[\frac{\hat{f}_{jt}(Z_{jt})}{f_j(Z_{jt})}\right] \\
& - \ln\left[\frac{\hat{g}_t(X_t)}{g(X_t)}\right] - \ln\left[\frac{\hat{g}_{t-j}(X_{t-j})}{g(X_{t-j})}\right]\Bigg\} \\
\equiv& \hat{\mathcal{I}}_{jn}(f_j, g \circ g) + \hat{\mathcal{I}}_{jn}(\hat{f}_j, f_j) - \hat{\mathcal{I}}_{1jn}(\hat{g}, g) - \hat{\mathcal{I}}_{2jn}(\hat{g}, g).
\end{aligned}
$$

Heuristic on the Asymptotic Expansion of the Entropy Estimators (3/8)

Under $\mathbb{H}_0$ that $\{X_t\}$ is i.d.d., using the inequality

$$\forall |u| < 1 : \left| \ln(1+u) - u + \frac{u^2}{2} \right| \leq |u|^3,$$

we obtain

$$\hat{\mathcal{I}}_{jn}(\hat{f}_j, f_j) = \hat{W}_1(j) + \frac{1}{2}\hat{W}_2(j) + \text{remainder},$$

where

$$\hat{W}_1(j) \equiv \frac{1}{n_j} \sum_{t=j+1}^{n} \frac{\hat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})}$$

$$\hat{W}_2(j) \equiv \frac{1}{n_j} \sum_{t=j+1}^{n} \left[\frac{\hat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right]^2.$$

Pairs Trading Strategies | Non-Parametric Entropy Measures | Testing the i.i.d. Hypothesis | References
Heuristic on the Asymptotic Expansion of the Entropy Estimators

Heuristic on the Asymptotic Expansion of the Entropy Estimators (4/8)

The first-order term $\hat{W}_1(j)$ of our Taylor series expansion can be approximate as

$$\hat{W}_1(j) = \frac{1}{2}\hat{H}_{1n}(j) + \text{remainder},$$

where $\hat{H}_{1n}(j)$ is a second-order U-statistic arising from the interaction between the sampling variation of the estimator $\hat{f}_{jt}(\cdot)$ and the averaging operation over the sample in $\hat{W}_1(j)$.

Heuristic on the Asymptotic Expansion of the Entropy Estimators (5/8)

The second-order term $\hat{W}_2(j)$ of our Taylor series expansion can be approximate as

$$\begin{aligned}\hat{W}_2(j) = & \frac{1}{n_j}\sum_{t=j+1}^{n}\int \mathrm{E}\left[\frac{\hat{f}_{jt}(z)-f_j(z)}{f_j(z)}\right]^2 f_j(z)\,\mathrm{d}\,z \\ & + \frac{1}{n_j}\sum_{t=j+1}^{n}\int\left\{\left[\frac{\hat{f}_{jt}(z)-f_j(z)}{f_j(z)}\right]^2 - \mathrm{E}\left[\frac{\hat{f}_{jt}(z)-f_j(z)}{f_j(z)}\right]^2\right\} f_j(z)\,\mathrm{d}\,z \\ & + \text{remainder} \\ = & \; L_n(j) + \hat{H}_{2n}(j) + \text{remainder},\end{aligned}$$

where $L_n(j)$ is the integrated weighted MSE of $\hat{f}_{jt}(\cdot)$, and $\hat{H}_{2n}(j)$ is a second term U-statistic.

Heuristic on the Asymptotic Expansion of the Entropy Estimators (6/8)

Therefore we obtain

$$2\hat{\mathcal{I}}_n(j) = -\frac{d_n^0}{n_j} + \hat{H}_n(j) + \text{remainder}$$

under $\mathbb{H}_0$, where $\hat{H}_n(j) \equiv \hat{H}_{1n}(j) - \hat{H}_{2n}(j)$ and the non-stochastic factor $d_n^0 = (A_n^0 - 1)^2$ and

$$A_n^0 \equiv \left(\frac{1}{h} - 2\right)\int_{-1}^{1} k^2(u)\,\mathrm{d}\,u + 2\int_0^1\int_{-1}^{b} k_b^2(u)\,\mathrm{d}\,u\,\mathrm{d}\,b,$$

where k_b is the jackknife kernel.

Heuristic on the Asymptotic Expansion of the Entropy Estimators (7/8)

Theorem 3.1

Suppose Assumptions A.1 and A.2 and $\mathbb{H}_0$ hold, $nh_n^4/\ln n \to \infty$, and $nh_n^7 \to 0$.

- Then $2h_n n_j \hat{\mathcal{I}}(j) + hd_n^0 \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ for any lag order $j = o(n)$, where

$$\sigma^2 \equiv 2 \int_{-1}^{1} \int_{-1}^{1} \Big[2k(u)k(u') - \int_{-1}^{1} k(u+v)k(v)\,\mathrm{d}\,v \times \int_{-1}^{1} k(u'+v')k(v')\,\mathrm{d}\,v' \Big]^2 \mathrm{d}\,u\,\mathrm{d}\,u'.$$

Pairs Trading Strategies | Non-Parametric Entropy Measures | Testing the i.i.d. Hypothesis | References
Heuristic on the Asymptotic Expansion of the Entropy Estimators

Heuristic on the Asymptotic Expansion of the Entropy Estimators (8/8)

Theorem 3.1–Continued

Suppose Assumptions A.1 and A.2 and $\mathbb{H}_0$ hold, $nh_n^4/\ln n \to \infty$, and $nh_n^7 \to 0$.

- Put $\hat{\mathcal{I}}_n \equiv [2hn_1\hat{\mathcal{I}}_n(1) + hd_n^0, \ldots, 2hn_p\hat{\mathcal{I}}_n(p) + hd_n^0]'$, where $p \in \mathbb{N}^+$ is fixed but may be arbitrarily large. Then $\hat{\mathcal{I}}_n \xrightarrow{d} \mathcal{N}(\vec{0}, \sigma^2 I_p)$, where I_p is a $p \times p$ identity matrix.

We first consider a test for $\mathbb{H}_0$ that is based on an individual lag j:

$$\mathcal{T}_n(j) \equiv \sigma^{-1}\Big[2h_n n_j \hat{\mathcal{I}}_n(j) + hd_n^0\Big], \; j \ll n.$$

By this Theorem, $\mathcal{T}_n(j) \xrightarrow{d} \mathcal{N}(0,1)$ under $\mathbb{H}_0$.

Testing the i.i.d. Hypothesis (1/5)

However,

- $\mathcal{T}_n(j)$ is a large sample test,
- its finite sample level may differ substantially from the asymptotic level.

The quality of the asymptotic approximation depends on

- the higher-order terms of the Taylor series expansion for $\hat{\mathcal{I}}_n(j)$,
- the choice of the bandwidth h.

Fortunately, because $\{X_t\}$ is i.i.d. under $\mathbb{H}_0$, the bootstrap is well suited and provides a simple way to obtain reasonable critical values for $\mathcal{T}_n(j)$.

Testing the i.i.d. Hypothesis (2/5)

Hong and White (2005) propose the following smoothed bootstrap procedure:

1. Draw a bootstrap sample $\mathcal{X}^* \equiv \{\mathcal{X}_t^*\}_{t=1}^n$ from the smoothed kernel density
$$\hat{g}(x) \equiv \frac{1}{n}\sum_{t=1}^{n} K_h(x, X_t), \; x \in \mathbb{I},$$
where $k(\cdot)$ and h are the same as those used in $\hat{\mathcal{I}}_n(j)$.
2. Compute a bootstrap entropy statistic $\hat{\mathcal{I}}_n^*(j)$ in the same way as $\hat{\mathcal{I}}_n(j)$, with $\mathcal{X}^*$ replacing $\mathcal{X}$.
3. Repeat steps (1) and (2) B times to obtain B bootstrap test statistic $\{\hat{\mathcal{I}}_{n\ell}^*(j)\}_{\ell=1}^B$.
4. Compute the bootstrap p-value
$$p^* \equiv B^{-1}\sum_{\ell=1}^{B} \mathbb{1}\left[\hat{\mathcal{I}}_{n\ell}^* > \hat{\mathcal{I}}_n(j)\right].$$

Testing the i.i.d. Hypothesis (3/5)

To show that the smoothed bootstrap procedure is consistent, we impose the following additional conditions on $k(\cdot)$.

Assumption A.3

Suppose $k : [-1, 1] \to \mathbb{R}^+$ is twice continuously differentiable on $[-1, 1]$ with $k^{(d)}(-1) = k^{(d)}(1) = 0$ for $d = 0, 1$ and $|k^{(2)}(u_1) - k^{(2)}(u_2)| \leq C|u_1 - u_2|$ for $u_1, u_2 \in [-1, 1]$.

Theorem: 4.1

Suppose Assumption A.1-A.3 and $\mathbb{H}_0$ hold, $nh^5 = O(1)$, $nh^7 \ln^3 n \to 0$, and $j = o(n)$. Let $\mathcal{T}_n^*$ be defined as $\mathcal{T}_n(j)$, with the bootstrap sample $\mathcal{X}^*$ defined above replacing the original sample $\mathcal{X}$ and with the same $k(\cdot)$ and h used in $\mathcal{T}_n(j)$, $\mathcal{T}_n^*(j)$, and $\hat{g}(\cdot)$. Then $\mathcal{T}_n^*(j) \xrightarrow{d} \mathcal{N}(0, 1)$ conditional on $\mathcal{X}$.

Pairs Trading Strategies | Non-Parametric Entropy Measures | Testing the i.i.d. Hypothesis | References
Testing the i.i.d. Hypothesis

Testing the i.i.d. Hypothesis (4/5)

It is thus desirable to have a portmanteau test that uses multiple lags. For this purpose, we consider

$$\mathcal{Q}_n(p) \equiv \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \mathcal{T}_n(j), \quad p \in \mathbb{N}^+.$$

For simplicity, we consider a fixed lag truncation number $p \in \mathbb{N}^+$.

Testing the i.i.d. Hypothesis (5/5)

Theorem 4.3 (i)

Suppose the conditions of Theorem 4.1 and $\mathbb{H}_0$ hold. Let $\mathcal{Q}_n^*(p)$ be defined as $\mathcal{Q}_n(p)$, with the bootstrap sample $\mathcal{X}^*$ replacing the original sample $\mathcal{X}$ but with the same $k(\cdot)$ and h used in $\mathcal{Q}_n(p)$, $\mathcal{Q}_n^*(p)$, and $\hat{g}(\cdot)$. Then for any fixed $p \in \mathbb{N}^+$, $\mathcal{Q}_n(p) \xrightarrow{d} \mathcal{N}(0,1)$ and $\mathcal{Q}_n^*(p) \xrightarrow{d} \mathcal{N}(0,1)$ conditional on $\mathcal{X}$.

Testing the i.i.d. $U[0,1]$ Hypothesis (1/4)

Under $\mathbb{H}_0^U$ that i.i.d. $U[0,1]$ hypothesis, and then the test statistic

$$\mathcal{T}_n^U \equiv \sigma^{-1}\left[2h \sum_{t \in S_n(j)} \ln \hat{f}_{jt}(Z_{jt}) + h\left[(A_n^0)^2 - 1\right]\right].$$

When multiple lags are considered, we can defined

$$\mathcal{Q}_n^U(p) \equiv \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \mathcal{T}_n^U(j), \ \ p \in \mathbb{N}^+.$$

Testing the i.i.d. $U[0,1]$ Hypothesis (2/4)

For this special hypothesis $\mathbb{H}_0^U$, Assumption A.1 and A.3 can be replaced by the following Assumption A.4.

Assumption A.4

Assume $\{X_t\}$ is a strictly stationary α-mixing process with mixing coefficient $\alpha(j) \leq Cj^{-\nu}$ for some $\nu > 2$. For each $j \in \mathbb{N}^+$, the joint probability density $f_j(\cdot)$ of Z_{jt} has support $\mathbb{I}^2$, is bounded away from 0, and is twice continuously differentiable on $\mathbb{I}^2$.

Testing the i.i.d. $U[0,1]$ Hypothesis (3/4)

Theorem 4.4

Suppose Assumptions A.2 and A.4, and the hypothesis $\mathbb{H}_0^U$ that $\{X_t\} \sim$ i.i.d. $U[0,1]$ hold and $nh^5 = O(1)$, $h \to 0$. Let $\mathcal{T}_n^U(j)^*$ and $\mathcal{Q}_n^U(p)^*$ be defined as $\mathcal{T}_n^U(j)$ and $\mathcal{Q}_n^U(p)$, respectively, with the bootstrap sample $\mathcal{X}^*$ replacing the original sample $\mathcal{X}$, where $\mathcal{X}^*$ is an i.i.d. sample drawn from the $U[0,1]$ distribution. The same $k(\cdot)$ and h are used in all test statistic. Then

- for any lag order $j = o(n)$, $\mathcal{T}_n^U(j) \xrightarrow{d} \mathcal{N}(0,1)$ and $\mathcal{T}_n^U(j)^* \xrightarrow{d} \mathcal{N}(0,1)$ conditional on $\mathcal{X}$;
- for any fixed $p \in \mathbb{N}^+$, $\mathcal{Q}_n^U \xrightarrow{d} \mathcal{N}(0,1)$ and $\mathcal{Q}_n^U(p)^* \sim \mathcal{N}(0,1)$ conditional on $\mathcal{X}$.

Testing the i.i.d. $U[0,1]$ Hypothesis (4/4)

The following theorem show there is a transformation from $\beta' y_t$ to $F(X)$.

Theorem: the Probability Integral Transform

Suppose X has a continuous distribution function F. Then $F(X)$ is uniformly distributed on $(0,1)$, and the probability integral transformation is the transformation from X to $F(X)$.

Since the strictly stationary and normal property of process $\beta' y_t$ implies Assumption A.2 and A.4,

We can test our pairs trading strategies!!!

Thank you for your time.

References (1/2)

[1] R. J. Elliott, J. Van Der Hoek, and W. P. Malcolm. Pairs trading. *Quantitative Finance*, 5(3):271–276, jun 2005.

[2] E. Gatev, W. N. Goetzmann, and K. G. Rouwenhorst. Pairs trading: performance of a relative-value arbitrage rule. *Review of Financial Studies*, 19(3):797–827, 2006.

[3] C. Granger and J.-L. Lin. Using the mutual information coefficient to identify lags in nonlinear models. *Journal of Time Series Analysis*, 15(4):371–384, jul 1994.

[4] Y. Hong and H. White. Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica*, 73(3):837–901, may 2005.

References (2/2)

[5] J. W. Jurek and H. Yang. Dynamic portfolio selection in arbitrage. *SSRN Electronic Journal*, 2007.

[6] E. S. Shreve. *Stochastic Calculus for Finance II*. Springer-Verlag New York, New York, 1 edition, 2004.

[7] G. Vidyamurthy. *Pairs Trading: Quantitative Methods and Analysis*. John Wiley & Sons, Inc., New Jersey, 2004.