TESTING NEW PROPERTY OF ELLIPTICAL MODEL FOR STOCK RETURNS DISTRIBUTION

Petr Koldanov

Joint work with Dmitriy Semenov
National Research University Higher School of Economics,
Moscow, Russia
pkoldanov@hse.ru

Yandex, Moscow, Russia, May 19-20, 2019

Outline

Introduction

- Models of multivariate distributions of stock returns
- Distribution models - portfolio selection and risk management.
- Popular model - elliptical distribution(Gupta et al. (2013)).
- Consistency of real data with elliptically contoured model was studied in (Chicheportiche and Bouchaud (2012))
- joint distribution of real market stock returns is not in accordance with hypothesis of elliptical distributions.
- Comparison of dependence measures of pairs of stock returns.
- Methodology in (Chicheportiche and Bouchaud (2012)) differs from usual hypothesis testing using statistical tools.
- Complement the analysis in (Chicheportiche and Bouchaud (2012)) by multiple hypotheses testing methodology.

Introduction

- Large stock market - the number of pairs of stocks is huge - multiplicity phenomenon (Bretz et al.(2011)).
- Multiple comparison procedures - adjust statistical inferences - enable better decision making.
- One possible approach - sign symmetry properties of elliptical model for bivariate distribution of each pairs of stock returns - was studied in (Koldanov and Lozgacheva (2016)).
- Main goal - detect pairs of stocks for which sign symmetry hypotheses are rejected - study associated rejection graph.
- Holm procedure (Holm (1979)) for multiple testing to detect non elliptical pairs of stocks.
- Holm procedure - strong control of FWER (probability of at least one Type I error).

Introduction

- New property of elliptical model - property of equality of $\tau-$Kendall correlation and probability of sign coincidence.
- Approach - testing new property for bivariate distribution of each pairs of stock returns - is studied.
- For concrete year - main goal - detect pairs of stocks for which hypotheses of new property are rejected - study associated rejection graph.
- Holm procedure (Holm (1979)) for multiple testing to detect non elliptical pairs of stocks.
- New procedure for combination of results of these Holm procedures is proposed.

Elliptical distribution

Random vector $X = (X_1, X_2, \dots, X_N)$ belongs to the class of elliptically contoured distributions if its density functions is (Anderson 2003):

$$f(x; \mu, \Lambda) = |\Lambda|^{-\frac{1}{2}} g\{(x-\mu)'\Lambda^{-1}(x-\mu)\} \qquad (1)$$

where Λ is positive definite matrix, $\mu = (\mu_1, \mu_2, \dots, \mu_N)$, $E(X_i) = \mu_i$, $g(x) \geq 0$, and

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(y'y) dy_1 \dots dy_N = 1$$

This class includes in particular multivariate Gaussian and multivariate Student distributions.

τ-Kendall correlation

Let $\begin{pmatrix} X_i \\ X_j \end{pmatrix}$ be random vector with cumulative distribution function $F_{X_iX_j}(x,y)$ and let $\begin{pmatrix} X_i(t) \\ X_j(t) \end{pmatrix}$, $\begin{pmatrix} X_i(t+1) \\ X_j(t+1) \end{pmatrix}$ be the independent copies of the vector $\begin{pmatrix} X_i \\ X_j \end{pmatrix}$.

Definition

Kendall measure $\tau_{i,j}$ of dependence between random variables X_i and X_j is defined by equality

$$\tau_{i,j} = P\left[(X_i(t+1) - X_i(t))(X_j(t+1) - X_j(t)) > 0\right]$$

Probability of sign coincidence[1]

Definition

Measure $\mathcal{Q}$ of dependence between random variables X_i and X_j is $\mathcal{Q}_{i,j} = P\left[(X_i - med(X_i))(X_j - med(X_j)) > 0\right]$ where $med(X_i)$ is the median of the distribution $F_{X_i}(x)$ of the random variable X_i i.e. $P(X_i > med(X_i)) = P(X_i < med(X_i)) = \frac{1}{2}$ or $F_{X_i}(med(X_i)) = \frac{1}{2}$.

It was proved that if random vector $X = (X_1, X_2, \dots, X_p)$ has multivariate normal distribution $N(\mu, \Lambda)$ with known μ then one has

$$\mathcal{Q}_{i,j} = \tau_{i,j} \text{ for all } i, j = 1, \dots, p; i \neq j \tag{2}$$

[1] William H. Kruskal (1958) Ordinal Measures of Association, Journal of the American Statistical Association, 53:284, 814-861

Theoretical results. Known μ

Let us prove that equality $\mathcal{Q}_{i,j} = \tau_{i,j}$ is true for elliptically contoured distributions $ECD(\mu, \Lambda, g)$ with any function g.

Theorem

If random vector X has elliptically contoured distribution $ECD(\mu, \Lambda, g)$ with known μ then for any g one has $\mathcal{Q}_{i,j} = \tau_{i,j}, \forall i, j = 1, \ldots, p, i \neq j$.

The theorem has the following

Corollary

If vector μ is known and exists $i, j = 1, \ldots, p; i \neq j$ such that $\mathcal{Q}_{i,j} \neq \tau_{i,j}$ then the vector X has not elliptically contoured distribution.

Theoretical results. Unknown μ

In real practice it is unrealistic to assume known vector μ. To deal with unknown μ let us prove the following

Theorem

Let $\begin{pmatrix} X_1(1) \\ \dots \\ X_p(1) \end{pmatrix}, \dots, \begin{pmatrix} X_1(n) \\ \dots \\ X_p(n) \end{pmatrix}$ *be the sample of independent identically distributed observations from vector* X *with elliptically contoured distribution* $ECD(\mu, \Lambda, g)$. *Let* $\overline{X_i} = \frac{1}{n}\sum_{t=1}^{n} X_i(t)$ *Then for* $n \geq 2$ *one has*

$$P((X_i(t) - \overline{X_i}) > 0, (X_j(t) - \overline{X_j}) > 0) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \frac{\lambda_{ij}}{\sqrt{\lambda_{ii}\lambda_{jj}}}$$

$, \forall i, j = 1, \dots, p; \forall t = 1, \dots, n$

Theoretical results. Unknown μ

It follows from the theorems that if vector X has elliptically contoured distribution $ECD(\mu, \Lambda, g)$ with any vector μ (known or unknown) then

$$P((X_i(t) - \overline{X_i})(X_j(t) - \overline{X_j}) > 0) = P((X_i(t) - \mu_i)(X_j(t) - \mu_j) > 0)$$

Since for $X - ECD(\mu, \Lambda, g)$ one has $med(X_i) = \mu_i$ then

$$P((X_i(t) - \overline{X_i})(X_j(t) - \overline{X_j}) > 0) = \mathcal{Q}_{i,j}$$

Corollary

If exists $i, j = 1, \ldots, p; i \neq j$ such that $\mathcal{Q}_{i,j} \neq \tau_{i,j}$ then the vector X has not elliptically contoured distribution.

Multiple testing framework

For study consistency with elliptical model it is necessary to test the equality $\mathcal{Q}_{i,j} = \tau_{i,j}$ for any $i, j = 1, \ldots, p; i \neq j$.
The τ-Kendall and probability of sign coincidence equality as individual hypotheses:

$$h_{i,j} : \mathcal{Q}_{i,j} = \tau_{i,j} \text{ against } k_{i,j} : \mathcal{Q}_{i,j} \neq \tau_{i,j} \quad (3)$$

- Main goal - detect pairs of stocks for which a hypotheses $h_{i,j}$ are rejected.
- Problem of simultaneous testing of individual hypotheses.
- The easiest approach - simply test each hypothesis at fixed level.
- The probability of one or more false rejections rapidly increases with number of hypotheses.
- In multiple testing - the requirement that the probability of one or more false rejections not exceed a given level.
- One such procedure was proposed in (Holm (1979))

Holm procedure

Let $p_{i,j}$ be p-value of the individual test for testing hypothesis $h_{i,j}$, $i,j = 1,2,\dots,N$, $i \neq j$, $M = \frac{N(N-1)}{2}$.

- Step 1: If $\min_{i,j=1,\dots,N} p_{i,j} \geq \frac{\alpha}{M}$ accept all hypotheses $h_{i,j}$, else, if $\min_{i,j=1,\dots,N} p_{i,j} = p_{i_1,j_1}$ reject hypothesis h_{i_1,j_1} and go to step 2.
- Step 2: Let $I = \{(i_1,j_1)\}$ be the set of indexes of previously rejected hypotheses. If $\min_{(i,j)\notin I} p_{i,j} \geq \frac{\alpha}{M-1}$ accept all hypotheses $h_{i,j}$, $(i,j) \notin I$, else if $\min_{(i,j)\notin I} p_{i,j} = p_{i_2,j_2}$ reject hypothesis h_{i_2,j_2} and go to step 3.
- ...
- Step M: Let $I = \{(i_1,j_1),\dots,(i_{M-1},j_{M-1})\}$ be the set of indexes of previously rejected hypotheses. Let $(i_M,j_M) \notin I$. If $p_{i_M,j_M} \geq \alpha$ accept the hypothesis h_{i_M,j_M}, else reject hypothesis h_{i_M,j_M} (reject all hypotheses).

The multiple testing procedure control FWER at level α.

τ-Kendall and probability of sign coincidence equality. Tests for individual hypotheses

Individual hypotheses:
$h_{i,j} : \mathcal{Q}_{i,j} = \tau_{i,j}$ vs $k_{i,j} : \mathcal{Q}_{i,j} \neq \tau_{i,j}$ $i, j = 1, \dots, N;\ i \neq j$
Without loss of generality let us consider the case $i = 1, j = 2$.
Let

$$\begin{pmatrix} x_1(1) \\ x_2(1) \end{pmatrix}, \begin{pmatrix} x_1(2) \\ x_2(2) \end{pmatrix}, \dots, \begin{pmatrix} x_1(n+m) \\ x_2(n+m) \end{pmatrix}$$

be the sample from random vector $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$.

τ-Kendall and probability of sign coincidence equality. Tests for individual hypotheses

For testing hypothesis $h_{1,2} : \mathcal{Q}_{1,2} = \tau_{1,2}$ against $k_{1,2} : \mathcal{Q}_{1,2} \neq \tau_{1,2}$ let us consider the statistics

$$\hat{\tau}_{1,2} = \sum_{i=1}^{[\frac{n}{2}]} S\{x_1(2i) - x_1(2i-1))(x_2(2i) - x_2(2i-1))\} \qquad (4)$$

and

$$\hat{\mathcal{Q}}_{1,2} = \sum_{i=n+1}^{n+m} S((x_1(i) - \overline{x_1}) \times (x_2(i) - \overline{x_2})) \qquad (5)$$

where

$$S(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$$\overline{x_k} = \frac{1}{m} \sum_{i=n+1}^{n+m} x_k(i), k = 1, 2$$

τ-Kendall and probability of sign coincidence equality. Tests for individual hypotheses

Statistic $\hat{\tau}_{1,2}$ has binomial distribution $b(r, \tau_{1,2})$ where $r = \left[\frac{n}{2}\right]$, statistic $\hat{\mathcal{Q}}_{1,2}$ has binomial distribution $b(m, \mathcal{Q}_{1,2})$. Moreover statistics $\hat{\tau}_{1,2}$, $\hat{\mathcal{Q}}_{1,2}$ are independent since they are based on different observations. Therefore

$$P(\hat{\tau}_{1,2} = k, \hat{\mathcal{Q}}_{1,2} = l) = C_r^k \tau_{1,2}^k (1 - \tau_{1,2})^{r-k} C_m^l \mathcal{Q}_{1,2}^l (1 - \mathcal{Q}_{1,2})^{m-l} =$$
$$= C_r^k C_m^l \exp\left\{ k \ln(\tfrac{\tau_{1,2}}{1-\tau_{1,2}}) + l \ln(\tfrac{\mathcal{Q}_{1,2}}{1-\mathcal{Q}_{1,2}}) \right\} (1 - \tau_{1,2})^r (1 - \mathcal{Q}_{1,2})^m$$

τ-Kendall and probability of sign coincidence equality. Tests for individual hypotheses

Then uniformly most powerful unbiased (UMPU) test for testing $h_{1,2} : \mathcal{Q}_{1,2} = \tau_{1,2}$ against $k_{1,2} : \mathcal{Q}_{1,2} \neq \tau_{1,2}$ has the form:

$$\varphi_{1,2}(x) = \begin{cases} 0, & c_1(l+k) \leq k \leq c_2(l+k) \\ 1, & k < c_1(l+k) \text{ or } k > c_2(l+k) \end{cases} \quad (6)$$

where $c_1(l+k), c_2(l+k)$ are defined from:

$$P_{\mathcal{Q}_{1,2}=\tau_{1,2}}(k < c_1(l+k) \text{ or } k > c_2(l+k)/\hat{\tau}_{1,2} + \hat{\mathcal{Q}}_{1,2} = l+k) = \alpha \quad (7)$$

For simplicity let us consider the test $\varphi_{1,2}(x)$ where constants $c_1(l+k), c_2(l+k)$ are defined from equations:

$$\begin{aligned} P_{\mathcal{Q}_{1,2}=\tau_{1,2}}(k < c_1(l+k)/\hat{\tau}_{1,2} + \hat{\mathcal{Q}}_{1,2} = l+k) = \tfrac{\alpha}{2} \\ P_{\mathcal{Q}_{1,2}=\tau_{1,2}}(k > c_2(l+k)/\hat{\tau}_{1,2} + \hat{\mathcal{Q}}_{1,2} = l+k) = \tfrac{\alpha}{2} \end{aligned} \quad (8)$$

P-values

Since

$$P_{\mathcal{Q}_{1,2}=\tau_{1,2}}(\hat{\tau}_{1,2}=k/\hat{\tau}_{1,2}+\hat{\mathcal{Q}}_{1,2}=l+k)=\frac{C_r^k C_m^l}{C_{r+m}^{k+l}}$$

then $c_1(l+k)$ is greatest integer number satisfying:

$$\sum_{i=0}^{c_1(l+k)}\frac{C_r^i C_m^{l+k-i}}{C_{r+m}^{k+l}}\leq\frac{\alpha}{2} \quad (9)$$

$c_2(l+k)$ is smallest integer number satisfying:

$$\sum_{i=c_2(l+k)}^{r}\frac{C_r^i C_m^{l+k-i}}{C_{r+m}^{k+l}}\leq\frac{\alpha}{2} \quad (10)$$

The p-value of the test (6) is defined as

$$q_{1,2}=2\min\left\{\sum_{i=0}^{k}\frac{C_r^i C_m^{l+k-i}}{C_{r+m}^{k+l}},\sum_{i=k}^{r}\frac{C_r^i C_m^{l+k-i}}{C_{r+m}^{k+l}}\right\} \quad (11)$$

Rejection graph

- Resulting distribution free Holm procedure was applied for 100 most traded stocks from Chinese stock market for 2006 year. The Holm procedure with FWER=0.05; 0.5 was applied to selected stocks for every year from the period from 2003 to 2014.
- To describe the results of Holm procedure the concept of *rejection graph* introduced in [2] was used.
- In all presented results the values $n = m = 125$ was selected. By the observations from the first half of a year the τ Kendall measure was estimated by the observations from the second half of a year the $\mathcal{Q}$ Kruscall measure was estimated.
- Rejection graph has unexpected structure. Despite number of rejected hypotheses the rejection graph has only few hubs and their removing leads to acceptance of all remaining symmetry hypotheses.

[2] Koldanov P., Lozgacheva N. MULTIPLE TESTING OF SIGN SYMMETRY FOR STOCK RETURN DISTRIBUTIONS // International Journal of Theoretical and Applied Finance. 2016. Vol. 19. No. 8. P. 1-14.

Rejection graphs for 2006, 2010 years

The obtained results are shown at the figures 1 (year 2006) and 2 (year 2010).

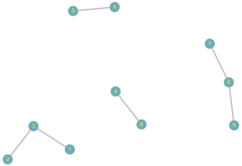

Figure: Rejection graph for Chinese market 2006 year. FWER=0.5. $n = m = 125$

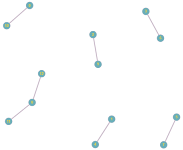

Figure: Rejection graph for Chinese market 2010 year. FWER=0.5. $n = m = 125$

New procedure for testing elliptical model

- Is it possible to accept the hypothesis of elliptical model of stock returns distribution for the period from 2003 to 2014 despite on the fact that for some years from the period the elliptical model was rejected by the Holm procedure?
- To answer this question, we propose the following procedure: the elliptical model is rejected if the number of years for which the elliptical model was rejected by the Holm procedure for one year period is greater than the given threshold.

Justification of the procedure

Let us formulate the hypothesis
H_i : 'in the year i the distribution of stock returns is elliptical' . Then the number X of rejected hypotheses $H_i, i = 1, \ldots, s$ for the period of length s is binomial random variable $b(s, \alpha)$ if all hypotheses $H_i, i = 1, \ldots, s$ are true and Holm procedure for each year has FWER$=\alpha$. Here independence of decisions obtained by the Holm procedure for one year period follows from independence of observations for different years.

New procedure

Let $H = \bigcap_{i=1}^{s} H_i$ be the hypothesis that elliptical model for stock returns distribution is correct for all considered years. Then

$$P_H(X \geq c) = \sum_{i=c}^{s} C_s^i (\alpha)^i (1-\alpha)^{s-i}$$

Therefore for given β one can define threshold c_β from the equation

$$P_H(X \geq c_\beta) = \sum_{i=c_\beta}^{s} C_s^i (\alpha)^i (1-\alpha)^{s-i} \leq \beta \tag{12}$$

and reject the hypothesis H iff $X \geq c_\beta$.

New procedure

Test of the hypothesis H has the form:

$$\varphi_H(x) = \begin{cases} 0, & x < c_\beta \\ 1, & x \geq c_\beta \end{cases} \tag{13}$$

where x is the observed number of the rejected hypotheses H_i, c_β is defined from (12) and β is the significance level of the test $\varphi_H(x)$. If the Holm procedure was applied with FWER$=\alpha$ then the p-value of the test (13) can be calculated from:

$$p_{H_\alpha}(x) = \sum_{i=x}^{s} C_s^i (\alpha)^i (1-\alpha)^{s-i} \tag{14}$$

New procedure. Experimental results

Rejected pairs from Chinese market for different years and for different α

year	$\alpha = 0.5$	$\alpha = 0.05$
2003	0	0
2004	0	0
2005	(88;23)	(88;23)
2006	(4;2)	(4;2)
2007	0	0
2008	0	0
2009	0	0
2010	(37;17),(85;66),(87;9),(92;56)	(37;17),(85;66),(87;9)
2011	(22;5),(16;10),(19;16)	(22;5),(16;10),(19;16)
2012	0	0
2013	(14;7),(29;2)	(14;7),(29;2)
2014	0	0

New procedure. Experimental results

For Chinese stock market the number of rejected hypotheses $h_{ij} : \mathcal{Q}_{i,j} = \tau_{i,j}$ is 11 for $\alpha = 0.5$ and 10 for $\alpha = 0.05$. From the other side the number of rejected hypotheses H_i is equal to 5 for $\alpha = 0.5$ and $\alpha = 0.05$ (for years 2005, 2006, 2010, 2011, 2013).
Then p-values (14) of the test $\varphi_H(x)$(13) of the hypothesis H are

$$p_{H_{0.5}}(5) = P(N \geq 5 | \alpha = 0.5) = \sum_{i=5}^{12} C_{12}^{i}(0.5)^{12} = 0.6128$$
$$p_{H_{0.05}}(5) = \sum_{i=5}^{12} C_{12}^{i}(0.05)^{i}(0.95)^{12-i} = 1.110779 * 10^{-5}$$

One can see that p-value $p_{H_{0.05}}(5)$ is too small. Then hypothesis H is rejected at any level $\beta > 1.110779 * 10^{-5}$ if we test individual hypotheses $H_i : i = 1, \dots, 12$ by Holm procedures with FWER 0.05. In contrast hypothesis H is accepted at any level $\beta \leq 0.6128$ if we test individual hypotheses $H_i : i = 1, \dots, 12$ by Holm procedures with FWER 0.5.

Conclusions

- New property of elliptically contoured distributions namely the property of equality of τ Kendall correlation coefficient and measure $\mathcal{Q}$ for any pair of random variables is proved in the paper.
- Distribution free individual tests for testing the property are constructed. Using well known Holm procedure these individual tests are combined to multiple hypotheses testing procedure.
- Application of the procedure to real market data from one year period shows that only small number of pairs of stocks from Chinese stock market destroy the tested property. Removing small number of stocks from consideration lead to nonrejection of elliptical model to the remaining stocks of stock market.
- New procedure of testing the elliptical model for the period of several years is proposed and applied for stock markets of China, USA, Great Britain and Germany for the period from 2003 to 2014. The obtained results shows that hypothesis of elliptical model is accepted for USA, Great Britain and Germany but is rejected for China.

Publications

- Gupta F.K. Varga T. Bodnar T. Elliptically Contoured Models in Statistics and Portfolio Theory, Springer, 2013.
- Chicheportiche R. Bouchaud J-P. The joint distribution of stock returns is not elliptical. International Journal of Theoretical and Applied Finance, 15, 3, 2012.
- Anderson T.W. An introduction to multivariate statistical analysis. Wiley-Interscience, New-York, 3-d edition, 2003.
- Bretz, Frank Hothorn, Torsten and Westfall, Peter. Multiple comparisons using R, Taylor and Francis Group, LLC, 2011.
- Holm, Sture. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics, 1979, 6 (2).
- Koldanov P., Lozgacheva N. Multiple testing of elliptical model for stock returns distribution. International Journal of Theoretic and Applied Finance, 2016. Vol. 19. No. 8. P. 1-14.

THANK YOU FOR YOUR ATTENTION!