Fundamental Analysis of Securities Trading
(IV) Pairs Trading - Test

Kuan-Lun WANG\textsuperscript{1,2}

National Taiwan University

Preliminary Draft: Friday 26\textsuperscript{th} June, 2020

\textsuperscript{1}E-mail: d06922002@csie.ntu.edu.tw
\textsuperscript{2}Home Page: https://www.csie.ntu.edu.tw/~d06922002/

---

Short Biodata

- **Research interests:**
  - time series models.
  - simulation modeling.
  - portfolio choice.
- **Central themes of my application:**
  - multivariate pairs trading in real time.
  - assets searching with a long-run equilibrium.
  - riskless portfolio building.
- **Current work:**
  - cointegration test.
  - structural change analysis.
  - the probability estimation of mean reversion.
Linearity Assumption (1/3)

We remark that

\[
\text{price}(t) = f(\text{factor}_1(t), \ldots, \text{factor}_n(t)) + \epsilon(t) \\
= a + \sum_{i=1}^{n} \sum_{j=1}^{\infty} b_{ij} \text{factor}_i^j(t) + \epsilon(t)
\]

and

\[
\begin{aligned}
\text{price}_A(t) &= a_A + b_A \text{factor}(t) + \epsilon_A(t) \\
\text{price}_B(t) &= a_B + b_B \text{factor}(t) + \epsilon_B(t).
\end{aligned}
\]

Linearity Assumption (2/3)

That is, we need to check the assumption. E.g.,

\[
\text{price}_A(t) = a + b \cdot \text{price}_B(t) + \text{something} + \epsilon(t) \\
= a + b \cdot \text{price}_B(t) + 0 + \epsilon(t).
\]

Consider a simple case \text{something} = t. i.e.,

\[
\text{price}_A(t) = a + b \cdot \text{price}_B(t) + c \cdot t + \epsilon(t).
\]
If \( \text{price}_{\text{portfolio}} \) with time linear trend, then one of technical indicator overbought and technical indicator oversold do not work. For example,
- for a positive time linear trend of portfolio, the oversold index maybe not work;
- for a negative time linear trend of portfolio, the oversold index maybe not work.

We can more safely use the data to reject than to confirm hypotheses [5].
- No swan is black (find one);
- Not all swans are white (check all).

In this case,
- \( H_0 \): All swans are white (0);
- \( H_1 \): No swan is black (1+).
If the strategy assumption is true, then earnings.
- \( H_0 \): Assumptions are true;
- \( H_1 \): One of assumptions is false.

<table>
<thead>
<tr>
<th>Decision</th>
<th>Not Reject ( H_0 )</th>
<th>Reject ( H_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0 ) is True</td>
<td>Negative</td>
<td>Positive</td>
</tr>
<tr>
<td>( H_0 ) is False</td>
<td>True Negative</td>
<td>False Positive</td>
</tr>
</tbody>
</table>

**Table:** The Probability Matrix for a Hypothesis Test

If the strategy assumption is true, then earnings.
Example: Testing of \( c \cdot t = 0 \)

We consider testing of

\[
H_0 : \text{price}_A(t) = a + b \cdot \text{price}_B(t) + 0 \cdot t + \epsilon(t)
\]

against

\[
H_1 : \text{price}_A(t) = a + b \cdot \text{price}_B(t) + c \cdot t + \epsilon(t).
\]

Decision Table: The Probability Matrix for a Hypothesis Test

<table>
<thead>
<tr>
<th>Decision</th>
<th>Not Reject ( H_0 )</th>
<th>Reject ( H_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0 ) is True</td>
<td>Correct ( (1 - \alpha) )</td>
<td>Type I Error ( (\alpha) )</td>
</tr>
<tr>
<td>( H_0 ) is False</td>
<td>Type II Error ( (\beta) )</td>
<td>Correct ( (1 - \beta) )</td>
</tr>
</tbody>
</table>

Hypothesis Notation [2]

- \( \alpha \): the significance level of the test;
- \( 1 - \alpha \): type I error; that is, rejecting \( H_0 \) and accepting alternative hypothesis \( H_1 \) when \( H_0 \) is true;
- \( \beta \): type II error; that is, failing to reject \( H_0 \) when \( H_1 \) is true;
- \( 1 - \beta \): power of the test.
Therefore, the economic consequence for the trading system is

$$\text{Profit} = (1 - \alpha) \cdot \text{Return}_{\text{Earning}} \cdot \Pr(H_0 \text{ is True}) + \beta \cdot \text{Return}_{\text{Losing}} \cdot \Pr(H_0 \text{ is False}).$$

In order to discuss the randomness, checking whether a coin is fair.

- **$H_0$:** Fair coin;
  - One for which the probability is 0.5.
- **$H_1$:** Unfair coin.
  - The heads probability < 0.5.

Let $X_1, \ldots, X_n$ be random variables associated with a flip coin trial by defining it as follows:

$$X_i(\text{heads}) \equiv 0 \text{ and } X_i(\text{tails}) \equiv 1 \text{ for all } i = 1, \ldots, n.$$
Setting the test significance level $\alpha \equiv 0.05$,

$$
\begin{align*}
\Pr \left( \sum_{i=1}^{n} X_i > \sum_{i=1}^{n} x_i \right) &< 1 - \alpha, \quad \text{reject } H_0; \\
\Pr \left( \sum_{i=1}^{n} X_i > \sum_{i=1}^{n} x_i \right) &\geq 1 - \alpha, \quad \text{not reject } H_0.
\end{align*}
$$
One-Tailed Test (4/6)

Figure: Rejection Region and Critical Value \((n = 100)\)

One-Tailed Test (5/6)

(a) Critical Experimental Probability

(b) Type II Error Rate \((n = 100 \text{ and } \alpha = 0.05)\)

Figure: One-Tailed Test
Figure: The $\alpha-\beta$ Trade-Off ($n = 100$ and $p = 0.55$)

- H$_0$: Fair coin; One for which the probability is 0.5.
- H$_1$: Unfair coin. The heads probability $\neq 0.5$.

Setting the test significance level $\alpha \equiv 0.05$,

$$
\begin{align*}
\text{Pr} \left( \sum_{i=1}^{n} x_i < \sum_{i=1}^{n} X_i < \sum_{i=1}^{n} x_i \right) &< 1 - \alpha, &\text{reject } H_0; \\
\text{Pr} \left( \sum_{i=1}^{n} x_i < \sum_{i=1}^{n} X_i < \sum_{i=1}^{n} x_i \right) &\geq 1 - \alpha, &\text{not reject } H_0.
\end{align*}
$$
Two-Tailed Test (2/4)

Figure: Rejection Region and Critical Value ($n = 100$)

Two-Tailed Test (3/4)

(a) Rejection Region Length

(b) Type II Error Rate ($n = 100$ and $\alpha = 0.05$)

Figure: Two-Tailed Test
Two-Tailed Test (4/4)

Figure: The $\alpha-\beta$ Trade-Off ($n = 100$ and $p = 0.55$)

Additional Restrictions (1/4)

Let

$$y \equiv \begin{pmatrix} \text{price}_A(1) \\ \vdots \\ \text{price}_A(T) \end{pmatrix}, \quad X = \begin{pmatrix} 1 & \text{price}_B(1) & 1 \\ \vdots & \vdots & \vdots \\ 1 & \text{price}_B(T) & T \end{pmatrix},$$

$$\beta \equiv \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \quad \text{and} \quad \epsilon \equiv \begin{pmatrix} \epsilon(1) \\ \vdots \\ \epsilon(t) \end{pmatrix}.$$ 

Then

$$y = X\beta + \epsilon.$$ 

In the sense that

$$\text{price}_A(t) = a + b \cdot \text{price}_B(t) + c \cdot t + \epsilon(t).$$
Additional Restrictions (2/4)

Let

\[
R \equiv \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad r \equiv 0.
\]

Then

\[
y = X\beta + \epsilon \quad \text{subject to} \quad R\beta = r,
\]

that meaning

\[
price_A(t) = a + b \cdot price_B(t) + 0 \cdot t + \epsilon(t).
\]

Additional Restrictions (3/4)

Additional Restrictions [1]

Under the OLS model, sometimes additional information about the unknown parameter vector \( \beta \) is available, which can be expressed in terms of a linear equation

\[
R\beta = r,
\]

where \( R \) is a known \( m \times p \) matrix of full row rank and \( r \) is a known \( m \times 1 \) vector.
Example: Testing of \( c \cdot t = 0 \)–Continued

We consider testing of

\[
H_0 : R\beta = 0
\]

against

\[
H_1 : R\beta \neq 0.
\]

Theorem: Lagrange Multiplier Theorem [4]

Let \( f : U \subseteq \mathbb{R}^n \to \mathbb{R} \) and \( g : U \subseteq \mathbb{R}^n \to \mathbb{R} \) be given \( C^1 \) functions. Let \( x_0 \in U \), \( g(x_0) = c_0 \), and let \( S = g^{-1}(c_0) \), the level set for \( g \) with value \( c_0 \). Assume \( \nabla g(x_0) \neq 0 \). If \( f|S \), which denotes \( f \) restricted to \( S \) (that is, to choose \( x \in U \) satisfying \( g(x) = c_0 \)), has a maximum or minimum at \( x_0 \), then there is a real number \( \lambda \) such that

\[
\nabla f(x_0) = \lambda \nabla g(x_0).
\]
Example: Maximize $x + y$ subject to $x^2 + y^2 = 1$

Let $U \equiv \mathbb{R}^2$, $f(x, y) \equiv x + y$, $g(x, y) \equiv x^2 + y^2$, $c_0 \equiv 1$, and $S \equiv g^{-1}(c_0) = g^{-1}(1)$. Since $(x, y) \in S$, $x \leq 1$ and $y = \pm \sqrt{1 - x^2}$, which implies that $f(x, y) = x \pm \sqrt{1 - x^2} \leq 2$. Then the upper bound for $f|S$ exists. By above theorem, there exists $\lambda \in \mathbb{R}$ such that

$$\nabla f(x, y)' = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} 2x \\ 2y \end{pmatrix} = \nabla g(x, y)' \implies x = y = \frac{1}{2\lambda}.$$ 

Since $(x, y) \in S$, $x^2 + y^2 = 1$, which implies that

$$\lambda = \pm \frac{1}{\sqrt{2}} \text{ and } (x, y) = \left( \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right), \left( -\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2} \right).$$

Example: Maximize $-xy$ subject to $x + y = 1$

Let $\lambda \equiv -1/2$ and $x_0, y_0 \equiv 1/2$. Then

$$\nabla f\left( \frac{1}{2}, \frac{1}{2} \right)' = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} = -\frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \nabla g\left( \frac{1}{2}, \frac{1}{2} \right)'$$

and

$$x_0 + y_0 = \frac{1}{2} + \frac{1}{2} = 1.$$ 

However,

$$(0)(1) = 0 > -\frac{1}{4} = -\left( \frac{1}{2} \right) \left( \frac{1}{2} \right) = -x_0y_0.$$
Lagrange Multiplier Theorem (4/4)

**Theorem: Lagrange Multiplier Theorem [4]**

Let \( f : U \subseteq \mathbb{R}^n \to \mathbb{R} \) and \( g : U \subseteq \mathbb{R}^n \to \mathbb{R}^m \) be given \( C^1 \) functions. Let \( x_0 \in U, \ g(x_0) = c_0 \), and let \( S = g^{-1}(c_0) \), the level set for \( g \) with value \( c_0 \). Assume \( \nabla g(x_0) \neq 0 \). If \( f|S \), which denotes \( f \) restricted to \( S \) (that is, to those \( x \in U \) satisfying \( g(x) = c_0 \)), has a maximum or minimum at \( x_0 \), then there is a \( \lambda \in \mathbb{R}^m \) such that

\[
\nabla f(x_0) = \lambda \nabla g(x_0) \implies (\nabla f(x_0))' = (\nabla g(x_0))' \lambda'.
\]

Restricted Least Squares Estimator (1/5)

Let \( R \in M_{m \times m}(\mathbb{R}) \) and \( r \in M_{m \times 1}(\mathbb{R}) \).

**Theorem: Restricted Least Squares Estimator**

Under the linear regression model with fundamental assumptions, the function

\[
f(\beta_*) = \|y - X\beta_*\|^2 = (y - X\beta_*)'(y - X\beta_*)
\]

is minimized with respect to \( \beta_*|_{R\beta_* = r} \) for \( \beta_* = \hat{\beta}_R \), where

\[
\hat{\beta}_R = \hat{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r).
\]

We remark \( \hat{\beta} = (X'X)^{-1}X'y \).
Testing

Restricted Least Squares Estimator (2/5)

Example

We remark that

\[
\text{Price}_A(t) = a + b \cdot \text{Price}_B(t) + c \cdot t + \epsilon(t)
\]

\[
\begin{align*}
1 &= 0 + 0.75 \cdot 2 + 0 \cdot 1 + -0.5 \\
33 &= 0 + 0.75 \cdot 44 + 0 \cdot 2 + 0 \\
308 &= 0 + 0.75 \cdot 410 + 0 \cdot 3 + 0.5
\end{align*}
\]

Then

\[
\beta = \begin{pmatrix}
\hat{a} \\
\hat{b} \\
\hat{c}
\end{pmatrix} = \begin{pmatrix}
-1.00 \\
0.75 \\
0.50
\end{pmatrix}
\quad \text{and} \quad
\beta_R = \begin{pmatrix}
\hat{a}_R \\
\hat{b}_R \\
\hat{c}_R
\end{pmatrix} = \begin{pmatrix}
-0.31 \\
0.75 \\
0.00
\end{pmatrix}.
\]

Restricted Least Squares Estimator (3/5)

Proof.

Since \( f(\beta_*) \geq 0 \), the lower bound for \( f \) exists. By Lagrange multiplier theorem, there exists \( \lambda \in \mathbb{R}^m \) such that

\[-2X'y + 2X'X\hat{\beta}_R = R'\lambda,\]

which implies that

\[
\hat{\beta}_R = (X'X)^{-1}X'y + \frac{1}{2}(X'X)^{-1}R'\lambda = \hat{\beta} + \frac{1}{2}(X'X)^{-1}R'\lambda.
\]
Restricted Least Squares Estimator (4/5)

**Proof—Continued.**

By the constraint,

$$ r = R\hat{\beta}_R = R\hat{\beta} + \frac{1}{2}R(X'X)^{-1}R'\lambda, $$

which implies that

$$ \lambda = -2[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r). $$

Then

$$ \hat{\beta}_R = \hat{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \equiv (I) + (II). $$

Restricted Least Squares Estimator (5/5)

**Theorem: Consistency**

Under the linear regression model with fundamental assumptions and additional restrictions, if $\lim_{n\to\infty} \frac{1}{n} X'X = Q$, where $Q$ is symmetric positive definite, then $\hat{\beta}_R$ is consistent for $\beta$.

**Rule: Properties of Convergence in Probability [3]**

Suppose $\{x_T\}$ and $\{y_T\}$ are $(K \times 1)$ random vectors, and let $c$ be a fixed $(K \times 1)$ vector. If $\lim plim x_T$ and $\lim plim y_T$ exist, then

- $\lim plim (x_T \pm y_T) = \lim plim x_T \pm \lim plim y_T$;
- $\lim plim (c'x_T) = c' \lim plim x_T$.

**Proof.**

Since $(I) \equiv \beta \xrightarrow{p} \beta$, $(II) \xrightarrow{p} 0$, which implies that $(II) + (II) \xrightarrow{p} \beta$. 
Restricted Least Squares Testing (1/8)

**Theorem: Restricted Least Squares Testing (F Distribution)**

Assume that $\epsilon \sim \mathcal{N}_n(0_n, \sigma^2 I_n)$. The corresponding test statistic is given as

$$F = \left( \frac{(\hat{\beta}_R - \hat{\beta})' X' X (\hat{\beta}_R - \hat{\beta})}{y'My} \right) \left( \frac{n - p}{m} \right), M = I_n - X(X'X)^{-1}X'.$$

Under the null hypothesis,

$$F \sim F_{m,n-p}.$$

Restricted Least Squares Testing (2/8)

**Definition: $\chi^2$ Distribution [1]**

Let $z \sim \mathcal{N}_m(\mu, I_m)$. Then the distribution of $z'z$ is called non-central $\chi^2$ distribution with $m$ degrees of freedom and non-centrality parameter $\lambda = \mu'\mu/2$, denoted by $z'z \sim \chi^2_{(m,\lambda)}$. In case $\lambda = 0$ the distribution is called central $\chi^2$ distribution with $m$ degrees of freedom, denote by $z'z \sim \chi^2_{(m)}$.

**Example**

Let $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(2,1)$. Then $X'X \sim \chi^2_{(1)}$ and $Y'Y \sim \chi^2_{(2,1)}$. 

<table>
<thead>
<tr>
<th>Testing</th>
<th>Linearity</th>
<th>Normality</th>
<th>Other</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISTP, CSIE, NTU</td>
<td>18/30</td>
<td>18/30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Proof.

We remark that \( \hat{\beta}_R = (X'X)^{-1}R(X'X)^{-1}R'\hat{\beta} - r \), \( E[\hat{\beta}] = \beta \), and \( R\beta = 0 \) under null hypothesis. Then

\[
E\left[ X(\hat{\beta}_R - \hat{\beta}) \right] = X E[\hat{\beta}_R - \hat{\beta}]
\]

\[
= X(X'X)^{-1}R(X'X)^{-1}R'\left( \hat{\beta} - r \right)
\]

\[
= X(X'X)^{-1}R(X'X)^{-1}R'\hat{\beta} - r
\]

\[
= X(X'X)^{-1}R(X'X)^{-1}R'\cdot 0
\]

\[
= 0.
\]

Proof–Continued.

We remark that \( \hat{\beta} = (X'X)^{-1}X'y \) and \( y \sim \mathcal{N}_n(X\beta, \sigma^2 I_n) \). So, we have

\[
X(\hat{\beta}_R - \hat{\beta}) = X(X'X)^{-1}R(X'X)^{-1}R'\hat{\beta} - r
\]

\[
= X(X'X)^{-1}R(X'X)^{-1}R'\hat{\beta} - r
\]

\[
= X(X'X)^{-1}R(X'X)^{-1}R'X'y - X(X'X)^{-1}R(X'X)^{-1}R'\hat{\beta} + r
\]

\[
= X(X'X)^{-1}R(X'X)^{-1}R' - X(X'X)^{-1}X'y + r
\]

Let \( Q \equiv X(X'X)^{-1}R(X'X)^{-1}R'X' \). Thus,

\[
X(\hat{\beta}_R - \hat{\beta}) = Q\epsilon = (\sigma Q)(\sigma^{-1} \epsilon).
\]
Theorem [1]

Let \( z \sim \mathcal{N}_m(\mu, I_m) \) and let \( A \) be an \( m \times m \) symmetric idempotent matrix. Then \( z'Az \sim \chi^2_{(a,\lambda)} \) with \( a = \text{rk}(A) \) and \( \lambda = \mu'A\mu/2 \).

Proof–Continued.

Since \( \epsilon \sim \mathcal{N}_n(0_n, \sigma^2 I_n) \), \( \sigma^{-1}\epsilon \sim \mathcal{N}_n(0_n, I_n) \). Since \( \text{rk}(\sigma^2 Q^2) = \text{rk}(Q) = m \) (Why?),

\[
(\hat{\beta}_R - \beta)'X'(\hat{\beta}_R - \beta) \sim \chi^2_m.
\]

Proof–Continued.

We remark that \( y = X\beta + \epsilon \sim \mathcal{N}_n(0_n, \sigma^2 I_n) \). It meaning that

\[
y - X\hat{\beta} = (I - X(X'X)^{-1}X')y = My
\]

\[
= (I - X(X'X)^{-1}X')\epsilon
\]

\[
= M\epsilon
\]

\[
= (\sigma M)(\sigma^{-1}\epsilon).
\]

Since \( \sigma^{-1}\epsilon \sim \mathcal{N}_n(0_n, I_n) \) and \( \text{rk}(\sigma M) = \text{rk}(M) = n - p \) (Why?),

\[
y'My \sim \chi^2_{n-p}.
\]
Restricted Least Squares Testing (7/8)

**Definition: F Distribution [1]**

Let $\chi^2_{(m,\lambda)}$ and $\chi^2_{(n,\delta)}$ be stochastically independent. Then the distribution of

$$X = \frac{\chi^2_{(m,\lambda)}/m}{\chi^2_{(n,\delta)}/n}$$

is called non-central F distribution with $(m, n)$ degrees of freedom and non-centrality parameters $(\lambda, \delta)$. If $\delta = \lambda = 0$, then the distribution of $X$ is called central $F$ distribution with $m$ and $n$ degrees of freedom, denoted by $X \sim F(m, n)$.

---

Restricted Least Squares Testing (8/8)

**Theorem [1]**

Let $z \sim \mathcal{N}_m(\mu, I_m)$. Let $A$ and $B$ be two $h \times m$ and $k \times m$ non-stochastic matrices. If $AB' = 0$, then $Az$ and $Bz$ are stochastically independent.

**Proof–Continued.**

Since $(\sigma Q)(\sigma M)' = \sigma^2 QM = 0$ (Why?), $Q\epsilon$ and $M\epsilon$ are independent, which implies $F \sim F(m, n-p)$.
Ramsey RESET (1/4)

One of Ramsey regression specification error test (Ramsey RESET) is simply the conventional $F$-test of the hypothesis

$$H_0 : y = X\beta + \epsilon$$

versus

$$H_1 : y = X\beta + Z\gamma + \epsilon$$

where

$$Z = [Z_{ij}]_{n \times m} = \left[ (X_i \hat{\beta})^j \right]_{n \times m}.$$

Ramsey RESET (2/4)

Example

Consider $t = 1, \ldots, 100$ and

$$\begin{cases}
\text{Factor}_1(t) = t; \\
\text{Factor}_2(t) = \sqrt{t}\sin(t); \\
\text{Price}_A(t) = 1 + 2\text{Factor}_1(t) - 5\text{Factor}_2(t) + \epsilon_A(t); \\
\text{Price}_B(t) = 3 + 4\text{Factor}_1(t) + 6\text{Factor}_2(t) + \epsilon_B(t).
\end{cases}$$

where

$$\begin{pmatrix}
\epsilon_A(t) \\
\epsilon_B(t)
\end{pmatrix} \sim \mathcal{N}_2\left(0, \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \right)$$

Under null hypothesis, $\hat{\alpha} \approx 15.10$ and $\hat{\beta} \approx 0.43.$
Example–Continued

(a) Asset A and Asset B
(b) Portfolio

Figure: Price Series and Portfolio Price Series

(a) Type II $p$-Value at $\alpha = 0.05$
(b) The $\alpha-\beta$ Trade Off

Figure: The Power of Ramsey RESET (1,000,000 Re-Sample Size)
### Skewness (1/3)

**Definition: Skewness Measure** [2]

One measure of the asymmetry of the distribution is skewness and is defined by

\[
\frac{E[(X - \mu)^3]}{E[(X - \mu)^2]^{3/2}} = \frac{E[(X - \mu)^3]}{\sigma^3}.
\]

When a distribution is symmetrical about the mean, the skewness is equal to zero.

**Proposition: Normality has Zero Skewness Measure** [2]

The normal distribution is asymmetry that immediately that the skewness of it is 0.

### Skewness (2/3)

**Figure: Measure Binomial Distribution with Skewness**

- **(a)** Probability Density Function
- **(b)** Skewness
Skewness (3/3)

(a) Probability Density Function

(b) Skewness

Figure: Measure Chi-Square Distribution with Skewness

Kurtosis (1/3)

Definition: Kurtosis Measure [3]

One measure of the tailedness of the distribution is Kurtosis and is defined by

$$\frac{E[(X - \mu)^4]}{E[(X - \mu)^2]^{4/2}} - 3 = \frac{E[(X - \mu)^4]}{\sigma^4} - 3.$$

The reason to subtract off 3 is that the normal distribution has zero Kurtosis measure.
Kurtosis (2/3)

(a) Probability Density Function  (b) Kurtosis

Figure: Measure Binomial Distribution with Skewness

Kurtosis (3/3)

(a) Probability Density Function  (b) Kurtosis

Figure: Measure Chi-Square Distribution with Kurtosis
Jarque–Bera Test (1/6)

Proposition: An Asymptotic Distribution about Normality [3]

Let $X_t$ be Gaussian white noise with nonsingular covariance matrix $\Sigma$ and expectation $\mu$, $X_t \sim \mathcal{N}_K(\mu, \Sigma)$. Let

$$\overline{X} \equiv \frac{1}{T} \sum_{t=1}^{T} X_t \quad \text{and} \quad \hat{\Sigma} \equiv \frac{1}{T-1} \sum_{t=1}^{T} (X_t - \overline{X})(X_t - \overline{X})'.$$

Let $Y_t \equiv (Y_{1t}, \ldots, Y_{Kt})' \equiv P^{-1}(X_t - \overline{X})$, where $P$ is a matrix satisfying $PP' = \hat{\Sigma}$. Let $b_1 \equiv (b_{11}, \ldots, b_{1K})'$, where $b_{1i} \equiv \sum_{t=1}^{T} Y_{it}^3 / T$ for $i = 1, \ldots, K$, and $b_2 \equiv (b_{21}, \ldots, b_{2K})'$, where $b_{2i} \equiv \sum_{t=1}^{T} Y_{it}^4 / T$ for $i = 1, \ldots, K$. Then

$$\sqrt{T} \begin{pmatrix} b_1 \\ b_2 - 3K \end{pmatrix} \mathcal{d} \rightarrow \mathcal{N}_{2K} \left( 0_{2K}, \begin{pmatrix} 6I_K & 0 \\ 0 & 24I_K \end{pmatrix} \right).$$

Jarque–Bera Test (2/6)

Remark: Chi-Square $\chi^2_{(\nu)}$ Distribution [1]

Let $X \sim \mathcal{N}_K(0_K, I_K)$. Then $X'X \sim \chi^2_{(n)}$.

Proposition: Jarque–Bera Test [3]

Take $S = T b_1' b_1 / 6$ and $K = T (b_2 - 3K)'(b_2 - 3K) / 24$. Then the asymptotic distributions of $S$ ans $K$ are

$$S \mathcal{d} \rightarrow \chi^2_{(K)} \quad \text{and} \quad K \mathcal{d} \rightarrow \chi^2_{(K)}.$$

Moreover, define the Jarque-Bera normality test statistic $J = S + K$, and the asymptotic distribution of $J$ is

$$J \mathcal{d} \rightarrow \chi^2_{(2K)}.$$
Jarque–Bera Test (3/6)

Proof.

\[
\sqrt{T} \left( \begin{bmatrix} b_1 \\ b_2 - 3K \end{bmatrix} \right) \xrightarrow{d} N \left( 0, \begin{bmatrix} 6I_K & 0 \\ 0 & 24I_K \end{bmatrix} \right)
\]

\[
\Rightarrow \sqrt{T} b_1 \xrightarrow{d} N(0, 6I_K)
\]

\[
\Rightarrow \sqrt{T} b_1 \xrightarrow{d} N(0, I_K)
\]

\[
\Rightarrow \left( \sqrt{T} b_1 \right) \left( \sqrt{T} b_1 \right) \xrightarrow{d} \chi^2_I(K)
\]

\[
\Rightarrow S = T b_1 / 6 \xrightarrow{d} \chi^2_I(K)
\]

Python 初階證券交易分析 台灣大學資訊系統訓練班

Jarque–Bera Test (4/6)

Proof–Continued.

\[
\sqrt{T} \left( \begin{bmatrix} b_1 \\ b_2 - 3K \end{bmatrix} \right) \xrightarrow{d} N \left( 0, \begin{bmatrix} 6I_K & 0 \\ 0 & 24I_K \end{bmatrix} \right)
\]

\[
\Rightarrow \sqrt{T} b_2 - 3K \xrightarrow{d} N(0, 24I_K)
\]

\[
\Rightarrow \sqrt{T} (b_2 - 3K) \xrightarrow{d} N(0, I_K)
\]

\[
\Rightarrow \left( \sqrt{T} (b_2 - 3K) \right) \left( \sqrt{T} (b_2 - 3K) \right) \xrightarrow{d} \chi^2_I(K)
\]

\[
\Rightarrow K = T (b_2 - 3K)'(b_2 - 3K)/24 \xrightarrow{d} \chi^2_I(K)
\]
Jarque–Bera Test (5/6)

**Theorem**

Let $X_1, \ldots, X_n$ be independent chi-square random variables with $r_1, \ldots, r_n$ degrees of freedom, respectively. Then the distribution of the random variable $Y = \sum_{i=1}^{n} X_n$ is $\chi^2(\sum_{i=1}^{n} r_n)$.

**Proof–Continued.**

Since $\text{Cov}[b_1, b_2 - 3K] = 0$, $b_1$ and $b_2 - 3K$ are independent, which implies that $S$ and $K$ are independent. By the above theorem,

$$J = S + K \overset{d}{\to} \chi^2(2K)$$

Jarque–Bera Test (6/6)

(a) Experimental Type I Error Rate  

(b) Experimental Error Rate

**Figure:** Jarque–Bera Test at $\alpha = 0.05$ and $K = 1$ (1,000,000 Re-Sample Size)
### Other Testing Model Assumptions

- Heteroscedasticity testing
- Auto-correlation testing
- Structural Change testing
- Outlier testing

### References (1/1)