Fundamental Analysis of Securities Trading
(IV) Pairs Trading B

Kuan-Lun WANG

National Taiwan University

Preliminary Draft: Thursday 14\textsuperscript{th} May, 2020

---

\textsuperscript{1}E-mail: polyphonicared@gmail.com
\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.
Definition: Transpose

The transpose of an \( m \times n \) matrix \( A \) is the \( n \times m \) matrix \( A' \), which is obtained from \( A \) by writing the rows of \( A \) as the columns of \( A' \).

Example

\[
A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \text{and} \quad A' = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}
\]

Example–Continued

We remark that Factor(\( t \)) : 1, 10, 100.

Let

\[
X \equiv \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 10 & 0 & 0 \\ 1 & 100 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 10 \\ 0 & 0 & 1 & 100 \end{pmatrix}
\]

Then

\[
X' \equiv \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 10 & 100 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 10 & 100 \end{pmatrix}
\]
Definition: Invertible Matrix [1]

If $A$ is a square matrix, and if a matrix $B$ of the same size can be found such that $AB = BA = I$, then $A$ is said to be invertible (or non-singular) and $B$ is called an inverse of $A$. If no such matrix $B$ can be found, then $A$ is said to be singular.

Example: Invertible Matrix [1]

Let

$$A = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}$$

Then

$$AB = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$$

$$BA = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$$

Thus, $A$ and $B$ are invertible and each is an inverse of the other.
Example–Continued

We remark that

\[
X = \begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 10 & 0 & 0 \\
1 & 100 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & 1 & 10 \\
0 & 0 & 1 & 100
\end{pmatrix} \quad \Rightarrow \quad X'X = \begin{pmatrix}
3 & 111 & 0 & 0 \\
111 & 10 & 101 & 0 & 0 \\
0 & 0 & 3 & 111 \\
0 & 0 & 111 & 10 & 101
\end{pmatrix}.
\]

Then

\[
(X'X)^{-1} \approx \begin{pmatrix}
0.56 & -0.01 & 0 & 0 \\
-0.01 & 0.00 & 0 & 0 \\
0 & 0 & 0.56 & -0.01 \\
0 & 0 & -0.01 & 0.00
\end{pmatrix}.
\]

Theorem: Exactly One Solution [1]

If \( A \) is an invertible \( n \times n \) matrix, then for each \( n \times 1 \) matrix \( b \), the system of equations \( Ax = b \) has exactly one solution, namely, \( x = A^{-1}b \).

Notation

If \( A \) is not an invertible, then the equation has not solutions or has at least two solutions.
### Solvable Equation System (7/10)

Consider

\[
\begin{pmatrix}
  y_1 \\
  \vdots \\
  y_T
\end{pmatrix}
= 
\begin{pmatrix}
  1 & x_1 \\
  \vdots & \vdots \\
  1 & x_T
\end{pmatrix}
\begin{pmatrix}
  a \\
  b
\end{pmatrix}
\]

\[\implies y = X\beta\]
\[\implies X'y = X'X\beta\]
\[\implies X'X\beta = X'y\]
\[\implies (X'X)^{-1}X'y = (X'X)^{-1}X'y\]
\[\implies \beta = (X'X)^{-1}X'y\]

### Solvable Equation System (8/10)

**Example—Continued**

We remark that \(a_A = 3, b_A = 3, a_B = 4, b_B = 4\),

\[
\text{Price}_A(t) : 1, 33, 308; \\
\text{Price}_B(t) : 2, 44, 410,
\]

and

\[
X =
\begin{pmatrix}
  1 & 1 & 0 & 0 \\
  1 & 10 & 0 & 0 \\
  1 & 100 & 0 & 0 \\
  0 & 0 & 1 & 1 \\
  0 & 0 & 1 & 10 \\
  0 & 0 & 1 & 100
\end{pmatrix}
\]
Example–Continued

Let
\[
\begin{pmatrix}
1 \\
33 \\
308 \\
2 \\
44 \\
410
\end{pmatrix}, \quad \begin{pmatrix}
3 \\
3 \\
4 \\
4
\end{pmatrix} \implies y \approx X\beta.
\]

Then
\[
\hat{\beta} \equiv (X'X)^{-1}X'y = \begin{pmatrix}
-0.06 \\
3.08 \\
0.33 \\
4.10
\end{pmatrix} \quad \text{and} \quad X\hat{\beta} = \begin{pmatrix}
3.027 \\
30.77 \\
308.20 \\
4.43 \\
41.32 \\
410.24
\end{pmatrix}.
\]

So,
\[
\begin{pmatrix}
\text{Price}_\text{portfolio}(1) \\
\text{Price}_\text{portfolio}(2) \\
\text{Price}_\text{portfolio}(3)
\end{pmatrix} = \begin{pmatrix}
\hat{b}_A\text{Price}_A(1) - \hat{b}_A\text{Price}_B(1) \\
\hat{b}_A\text{Price}_A(2) - \hat{b}_A\text{Price}_B(2) \\
\hat{b}_A\text{Price}_A(3) - \hat{b}_A\text{Price}_B(3)
\end{pmatrix} = \begin{pmatrix}
-2.07 \\
-0.36 \\
-1.33
\end{pmatrix}.
\]
Definition: Matrix Addition and Matrix Subtraction [1]

If $A$ and $B$ are matrices of the same size, then the **sum** $A + B$ is the matrix obtained by adding the entries of $B$ to the corresponding entries of $A$, and the **difference** $A - B$ is the matrix obtained by subtracting the entries of $B$ from the corresponding entries of $A$. Matrices of different size cannot be added or subtracted.

Example: Matrix Addition and Matrix Subtraction [1]

Consider the matrices

$$A = \begin{pmatrix} 2 & 1 & 0 & 3 \\ -1 & 0 & 2 & 4 \\ 4 & -2 & 7 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} -4 & 3 & 5 & 1 \\ 2 & 2 & 0 & -1 \\ 3 & 2 & -4 & 5 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}.$$

Then

$$A + B = \begin{pmatrix} -2 & 4 & 5 & 4 \\ 1 & 2 & 2 & 3 \\ 7 & 0 & 3 & 5 \end{pmatrix} \quad \text{and} \quad A - B = \begin{pmatrix} 6 & -2 & -5 & 2 \\ -3 & -2 & 2 & 5 \\ 1 & -4 & 11 & -5 \end{pmatrix}.$$

The expressions $A + C$, $B + C$, $A - C$, and $B - C$ are undefined.
However,

\[
\begin{bmatrix}
y_1 \\ \\ \\ y_T
\end{bmatrix} = \begin{bmatrix}
1 & x_1 \\ \\ \\ 1 & x_T
\end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_T \end{bmatrix}
\]

\[\Rightarrow y = X\beta + \epsilon\]

\[\Rightarrow X'y = X'X\beta + X'\epsilon\]

\[\Rightarrow (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'\epsilon\]

\[\Rightarrow \beta = (X'X)^{-1}X'y - (X'X)^{-1}X'\epsilon\]
Example

Consider the matrices

\[
A = \begin{pmatrix}
1 & 2 & 3 \\
1 & 4 & 5 \\
1 & 5 & 6
\end{pmatrix} = (1 \cdot 1) + (1 \cdot 2) \quad \text{and} \quad B = \begin{pmatrix}
1 & 2 & 3 \\
1 & 4 & 2 \\
1 & 5 & 7
\end{pmatrix}.
\]

Then \( A \) is not of full column rank and \( B \) is of full column rank.

Fundamental Assumptions

The ordinary linear regression model is described by the equation

\[
y = X\beta + \epsilon,
\]

where

- \( X \) is a non-stochastic \( n \times p \) matrix with \( p < n \);
- the matrix \( X \) has rank \( p \), i.e. \( X \) is of full column rank;
- the elements of the \( n \times 1 \) vector \( y \) are observable random vectors;
- the elements of the \( n \times 1 \) vector \( \epsilon \) are non-observable random variables such that \( E[\epsilon] = 0 \) and \( \text{Cov}[\epsilon] = \sigma^2 I_n \) with \( \sigma^2 > 0 \).

We will write \( \epsilon \sim (0, \sigma^2 I_n) \) for short.

The linear regression with fundamental assumptions is also called classical linear regression model.
Example

Consider the linear regression

\[
\begin{pmatrix}
\text{Price}(1) \\
\vdots \\
\text{Price}(T)
\end{pmatrix} = \begin{pmatrix} 1 & \text{Factor}_1(1) & \cdots & \text{Factor}_p(1) \\
\vdots & \vdots & \ddots & \vdots \\
1 & \text{Factor}_1(T) & \cdots & \text{Factor}_p(T)
\end{pmatrix} \begin{pmatrix} a \\
\vdots \\
b_p
\end{pmatrix} + \begin{pmatrix} \epsilon(1) \\
\vdots \\
\epsilon(T)
\end{pmatrix}.
\]

Example—Continued

We remark that \(a_A = 3\), \(b_A = 3\), and

\[
\begin{align*}
\text{Price}_A(t) & : 1, 33, 308; \\
\text{Factor}(t) & : 1, 10, 100.
\end{align*}
\]

Let

\[
\begin{pmatrix} 1 \\
33 \\
308 \\
\end{pmatrix} = \begin{pmatrix} 1 & 1 \\
1 & 10 \\
1 & 100
\end{pmatrix} \begin{pmatrix} 3 \\
\vdots \\
3
\end{pmatrix} + \begin{pmatrix} -5 \\
0 \\
5
\end{pmatrix}.
\]
Example–Continued

We remark that $a_B = 4$, $b_B = 4$, and

$$\begin{align*}
\text{Price}_A(t) & : 2, 44, 410; \\
\text{Factor}(t) & : 1, 10, 100.
\end{align*}$$

Let

$$\begin{pmatrix} 2 \\ 44 \\ 410 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 10 \\ 1 & 100 \end{pmatrix} \begin{pmatrix} 4 \end{pmatrix} + \begin{pmatrix} -6 \\ 0 \\ 6 \end{pmatrix}.$$
Explanation (5/6)

(a) Histogram Plot

(b) 2-D Line Plot

Figure: A Classical Linear Regression Model Example

Explanation (6/6)

Figure: This $X$ can include commodities prices, product prices, etc.
Ordinary Least-Squares Estimator (1/9)

**Notation: Unsolvable Equation System**

If we assume that the system of equations $y = X\beta_*$ is solvable with respect to $\beta_*$, then a solution $\beta^0$ clearly satisfies $\|y - X\beta^0\|^2 = 0$. On the other hand, when we assume that $y = X\beta_*$ is not solvable, then we can nonetheless determine a vector $\hat{\beta}$ such that

$$\|y - X\hat{\beta}\|^2 \leq \|y - X\beta_*\|^2$$

for every vector $\beta_* \in \mathbb{R}^p$.

Ordinary Least-Squares Estimator (2/9)

Under the linear regression model, we interested in inference about the unknown parameters $\beta \in \mathbb{R}^p$ and $\sigma^2 > 0$. Thus, the parameter space is given by $\Theta = \mathbb{R}^p \times (0, \infty)$.

**Definition: Least-Squares Solution**

A vector $\hat{\beta}$ is called least squares solution of $y = X\beta_*$ if

$$\|y - X\hat{\beta}\|^2 \leq \|y - X\beta_*\|^2$$

for every vector $\beta_* \in \mathbb{R}^p$.

If we consider $\epsilon_* = y - X\beta_*$ as the residual vector of the solution $\beta_*$, then the sum of squared residuals is minimized for $\beta_* = \hat{\beta}$, so that $\hat{\beta}$ has the smallest sum of squared residuals.
Example

We remark that

\[
Y = \begin{pmatrix}
\text{Price}_A(1) \\
\text{Price}_A(2) \\
\text{Price}_A(3) \\
\text{Price}_B(1) \\
\text{Price}_B(2) \\
\text{Price}_B(3)
\end{pmatrix}
= \begin{pmatrix} 1 \\ 33 \\ 308 \\ 2 \\ 44 \\ 410 \end{pmatrix}
\]

and

\[
X \equiv l_2 \otimes \begin{pmatrix}
1 & \text{Factor}(1) \\
1 & \text{Factor}(2) \\
1 & \text{Factor}(3)
\end{pmatrix}
= l_2 \otimes \begin{pmatrix} 1 & 1 \\ 1 & 10 \\ 1 & 100 \end{pmatrix}
= \begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 10 & 0 & 0 \\
1 & 100 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & 1 & 10 \\
0 & 0 & 1 & 100
\end{pmatrix}.
\]

\[\beta \equiv \begin{pmatrix}
a_A \\
b_A \\
a_B \\
b_B
\end{pmatrix}
= \begin{pmatrix} 3 \\ 3 \\ 4 \\ 4 \end{pmatrix},
\hat{\beta} \equiv \begin{pmatrix}
\hat{a}_A \\
\hat{b}_A \\
\hat{a}_B \\
\hat{b}_B
\end{pmatrix}
= \begin{pmatrix} -0.06 \\ 3.08 \\ 0.33 \\ 4.10 \end{pmatrix}.
\]
Example–Continued

Then

\[
\begin{pmatrix}
-2.03 \\
2.23 \\
-0.20 \\
-2.43 \\
2.68 \\
-0.24
\end{pmatrix},
\begin{pmatrix}
-5 \\
0 \\
5 \\
-6 \\
0 \\
6
\end{pmatrix}
\]

and

\[
\|y - X\hat{\beta}\|^2 = 22.26 \leq 122 = \|y - X\beta\|^2 = \|\epsilon\|^2.
\]
Theorem: Ordinary 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 for \( \beta^* = \hat{\beta} \), where \( \hat{\beta} = (X'X)^{-1}X'y \). Moreover, the vector \( \hat{\beta} \) is called ordinary least-squares estimator of \( \beta \).

Proposition: Chain Rule for Vector Differentiation [3]

Let \( \alpha \) and \( \beta \) be \((m \times 1)\) and \((n \times 1)\) vectors, respectively, and suppose \( h(\alpha) \) is \((p \times 1)\) and \( g(\beta) \) is \((m \times 1)\). Then, with \( \alpha = g(\beta) \),

\[
\frac{\partial h(g(\beta))}{\partial \beta'} = \frac{\partial h(\alpha)}{\partial \alpha'} \frac{\partial g(\beta)}{\partial \beta'}.
\]

Rules of Matrix Calculus [3]

- Let \( A \) be an \((m \times n)\) matrix and \( \beta \) be an \((n \times 1)\) vector. Then
  \[
  \frac{\partial A\beta}{\partial \beta'} = A \quad \text{and} \quad \frac{\partial \beta'A}{\partial \beta} = A'.
  \]
- Let \( A \) be \((m \times m)\) and \( \beta \) be \((m \times 1)\). Then
  \[
  \frac{\partial \beta'A\beta}{\partial \beta} = (A' + A)\beta \quad \text{and} \quad \frac{\partial \beta'A\beta}{\partial \beta'} = \beta'(A' + A).
  \]
Proof.

By the differentiating the function \( f(\beta^*) \) with respect to \( \beta^* \),

\[
\frac{\partial f(\beta^*)}{\partial \beta^*} = \frac{\partial (y - X\beta^*)'(y - X\beta^*)}{\partial \beta^*} = (X'X)'(I' + I')
\]

and \( \partial^2 f(\beta^*)/\partial \beta^*^2 = 2X'X \) is positive definite. The solution is given by \( \hat{\beta} = (X'X)^{-1}X'y \) if we put the right-hand side equal to 0 and solve for \( \beta^* \).

---

**Maxwell Likelihood Estimator**

**Theorem: Maximum Likelihood Estimator**

Under the linear regression model with fundamental assumptions and assumption \( \epsilon \sim \mathcal{N}(0, \sigma^2 I_n) \), the likelihood function

\[
L_T(\beta^*, \sigma^2) = \prod_{t=1}^T \Pr(y_t; X_t; \beta^*, \sigma^2) = \prod_{t=1}^T \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{(y_t - X_t\beta^*)^2}{2\sigma^2} \right\}
\]

is maximized for \( \beta^* = \tilde{\beta} \), where \( \tilde{\beta} = (X'X)^{-1}X'y \). Moreover, the vector \( \tilde{\beta} \) is called maximum likelihood estimator of \( \beta \).
Proof.

By the definition of likelihood function,

\[
\log L_T(\beta_*, \sigma^2) = -T \log \sigma \sqrt{2\pi} - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_t - X_t \beta_*)^2 \\
= -T \log \sigma \sqrt{2\pi} - \frac{(y - X\beta_*)'(y - X\beta_*)}{2\sigma^2} \\
= -T \log \sigma \sqrt{2\pi} - \|y - X\beta_*\|^2 / 2\sigma^2.
\]

Then the proof of ordinary least-squares estimator theorem implies this theorem.

Example

In order to discuss the distribution, let

\[
X(t) = [1, x(t)], \ x(t) \sim U[0, 1], \ \beta \equiv \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \ \epsilon(t) \sim \mathcal{N}(0, 1).
\]

Then

\[
\text{Figure: } y = X\beta + \epsilon
\]
Maximum Likelihood Estimator

Example–Continued

Figure: \( \log L_T(\beta, \sigma^2) \)

**Theorem: Unbiased for \( \beta \)**

Under the linear regression model with fundamental assumptions,

\[
E[\hat{\beta}] = \beta \quad \text{and} \quad \text{Cov}[\hat{\beta}] = \sigma^2 (X'X)^{-1}
\]

hold true for \( \hat{\beta} = (X'X)^{-1}X'y \).

**Proof.**

Since \( y = X\beta + \epsilon \) and \( \epsilon \sim (0, \sigma^2) \),

\[
E[\hat{\beta}] = E[(X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon] = \beta
\]

and

\[
\text{Cov}[\hat{\beta}] = E[\hat{\beta}\hat{\beta}^T - \beta\beta^T] = \sigma^2 (X'X)^{-1}.
\]
Example

Let

\[ X \equiv \begin{pmatrix} 1 & 1 \\ 1 & 10 \\ 1 & 100 \end{pmatrix}, \quad \beta \equiv \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \quad \epsilon \sim \mathcal{N}_3(0, I), \quad y \equiv X\beta + \epsilon. \]

We simulate \(10^6\) time series. Then

\[
\text{Mean} \left[ \hat{\beta} \right] = \begin{pmatrix} 3.00 \\ 3.00 \end{pmatrix}, \quad \text{Cov} \left[ \hat{\beta} \right] = \begin{pmatrix} 0.56 & -0.01 \\ -0.01 & 0.00 \end{pmatrix} \approx \sigma^2 (X'X)^{-1}.
\]

Example–Continued

Then

- (a) Color Histogram Bars by Height
- (b) A Bi-Variate Tiled Histogram

Figure: Bi-Variate Histogram Plot
Definition: Convergence in Probability

A series $Y_1, \ldots, Y_n, \ldots$ of random vectors converges in probability to a fixed $c$, if

$$\forall i, \lim_{n \to \infty} \Pr(|Y_{n,i} - c| > \epsilon) = 0$$

for every $\epsilon > 0$. The symbol plim denotes convergence in probability.

Definition: Consistent

An estimator $\hat{\beta}$ of $\beta$ is called consistent for $\beta$, if $\lim_{n \to \infty} \hat{\beta} = \beta$ holds true.

Theorem: Consistency

Under the linear regression model with fundamental assumptions, if $\lim_{n \to \infty} \frac{1}{n} X'X = Q$, where $Q$ is symmetric positive definite, then $\hat{\beta}$ is consistent for $\beta$. (We call this sufficient condition the asymptotic assumption.)

Proof.

Under the assumptions, $(X'X)^{-1} = O(n^{-1})$ and

$$\text{Cov} \left[ \hat{\beta} \right] = \sigma^2 (X'X)^{-1} = O(n^{-1}).$$

Since $E[\hat{\beta}] = \beta$, $\hat{\beta}$ is consistent for $\beta$. 
Example

Let

\[
X \equiv \begin{pmatrix}
1 & 1 \\
\vdots & \vdots \\
1 & 10^{50}
\end{pmatrix},
\beta \equiv \begin{pmatrix}
a \\
b
\end{pmatrix} = \begin{pmatrix}3 \\ 3\end{pmatrix}, \quad \epsilon \sim \mathcal{N}_{50}(0, 1),
y \equiv X\beta + \epsilon.
\]

Since \(10^t \cdot 10^t / t\) is divergence, \(X'X\) is also.

---

Property (7/9)

Example—Continued

(a) \(Q\)

(b) \(\|\beta - \hat{\beta}\|\)

**Figure**: Divergence
Example

Let

\[ x(t) \sim U[0, 1], \quad X(t) = (1 \quad x(t)) , \quad \epsilon \sim N_{50}(0, 1), \quad y = X\beta + \epsilon. \]

Since \( 1 \cdot 1/t \to 0 \), \( \lim_{T \to \infty} X'X/T \) exists. Then \( \hat{\beta} \overset{p}{\to} \beta. \)

Example–Continued

\( (a) \quad Q \quad (b) \quad ||\beta - \hat{\beta}|| \)

Figure: Convergence
References (1/1)

