Fundamental Analysis of Securities Trading
(IV) Pairs Trading A

Kuan-Lun WANG\textsuperscript{1,2}

National Taiwan University

Preliminary Draft: February 9, 2020

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

Short Biodata

\begin{itemize}
  \item Research interests:
    \begin{itemize}
      \item time series models.
      \item simulation modeling.
      \item portfolio choice.
    \end{itemize}
  \item Central themes of my application:
    \begin{itemize}
      \item multivariate pairs trading in real time.
      \item assets searching with a long-run equilibrium.
      \item riskless portfolio building.
    \end{itemize}
  \item Current work:
    \begin{itemize}
      \item cointegration test.
      \item structural change analysis.
      \item the probability estimation of mean reversion.
    \end{itemize}
\end{itemize}
Assumption (1/2)

Consider two asset prices has only one same principal factor. That is,

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

Assume \(\epsilon_A\) and \(\epsilon_B\) are stable.

Warning

Arbitrage pricing theory shows this assumption is not good. But why we assume the above? (Hint: see the first slide.)

https://en.wikipedia.org/wiki/Arbitrage_pricing_theory

Assumption (2/2)

Wiki: Arbitrage Pricing Theory

Risky asset returns are said to follow a factor intensity structure if they can be expressed as:

\[
r_j = a_j + b_{j1} F_1 + b_{j2} F_2 + \cdots + b_{jn} F_n + \epsilon_j
\]

where

- \(a_j\) is a constant for asset \(j\),
- \(F_n\) is a systematic factor,
- \(b_{jn}\) is the sensitivity of the \(j\)-th asset to factor \(n\), also called factor loading,
- and \(\epsilon_j\) is the risky asset’s idiosyncratic random shock with mean zero.
If we buy \( b_B \) A and short \( b_A \) B at time \( t_0 \), then our cash flow is

\[
\begin{align*}
&= -b_B \text{price}_A(t_0) + b_A \text{price}_B(t_0) \\
&= -b_B (a_A + b_A \text{factor}(t) + \epsilon_A(t)) \\
&\quad + b_A (a_B + b_B \text{factor}(t) + \epsilon_B(t)) \\
&= (-a_A b_B + a_B b_A) \\
&\quad + (-b_A b_B \text{factor}(t) + b_A b_B \text{factor}(t)) \\
&\quad + (-b_B \epsilon_A(t_0) + b_A \epsilon_B(t_0)) \\
&= (-a_A b_B + a_B b_A) + (-b_B \epsilon_A(t_0) + b_A \epsilon_B(t_0)).
\end{align*}
\]

In fact, our cash flow is

\[
- b_B \text{price}_A(t_0) - b_A \text{price}_B(t_0).
\]

Remark: Our Assumptions

The prices process in the following form

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

where \( \epsilon_A \) and \( \epsilon_B \) are stable.

That is, the portfolio value is

\[
\text{price}_{\text{portfolio}}(t) = (a_A b_B - a_B b_A) + (b_B \epsilon_A(t) - b_A \epsilon_B(t)).
\]

Moreover, the value is independent of \( \text{factor}(t) \).
So, if we know the values $b_A$ and $b_B$, then we can build a portfolio independent of principal factor. Similarly, we can consider any number of factor.

Why do we need to know this?
If we buy one this portfolio at time $t_0$ and sell it at time $t_1$, then our cash flow is

$$\begin{align*}
- \text{price}_{\text{portfolio}}(t_0) + \text{price}_{\text{portfolio}}(t_1) \\
= & -((a_A b_B - a_B b_A) + (b_B \epsilon_A(t_0) - b_A \epsilon_B(t_0))) \\
& + ((a_A b_B - a_B b_A) + (b_B \epsilon_A(t_1) - b_A \epsilon_B(t_1))) \\
= & -(a_A b_B - a_B b_A) + (a_A b_B - a_B b_A) \\
& + (- (b_B \epsilon_A(t_0) - b_A \epsilon_B(t_0)) + (b_B \epsilon_A(t_1) - b_A \epsilon_B(t_1))) \\
= & -(b_B \epsilon_A(t_0) - b_A \epsilon_B(t_0)) + (b_B \epsilon_A(t_1) - b_A \epsilon_B(t_1)).
\end{align*}$$

Remark: Return of This Portfolio

Return of this portfolio over $[t_0, t_1]$ is

$$\text{Return}_{[t_0, t_1]} = -(b_B \epsilon_A(t_0) - b_A \epsilon_B(t_0)) + (b_B \epsilon_A(t_1) - b_A \epsilon_B(t_1)).$$

That shows we only trade a stable noise. We can wait for the low value to buy it or wait for the high value to sell it.
Definition: Matrix

A rectangular arrangement

\[ A = \begin{pmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  a_{21} & a_{22} & \cdots & a_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{m1} & a_{m2} & \cdots & a_{mn}
\end{pmatrix} \]

of elements \( a_{ij}, \ i = 1, \ldots, m, \ j = 1, \ldots, n \), in \( m \times n \) rows and \( n \) columns is called an \( m \times n \) matrix \( A \).

Example: Matrix [1]

Some examples of matrices are

\[ \begin{pmatrix} 1 & 2 \\ 3 & 0 \\ -1 & 4 \end{pmatrix}, \begin{pmatrix} 2 & 1 & 0 & -3 \end{pmatrix}, \begin{pmatrix} e & \pi & -\sqrt{2} \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix}, (4). \]
**Definition: Matrix Product** [1]

If $A$ is an $m \times r$ matrix and $B$ is an $r \times n$ matrix, then the product $AB$ is the $m \times n$ matrix whose entries are determined as follows:

To find the entry in row $i$ and column of $AB$, single out row $i$ from the matrix $A$ and column $j$ from the matrix $B$. Multiple the corresponding entries from the row and column together, and then add up the resulting products.

**Example: Multiplying Matrices** [1]

Consider the matrices

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 6 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 4 & 1 & 4 & 3 \\ 0 & -1 & 3 & 1 \\ 2 & 7 & 5 & 2 \end{pmatrix}$$

Since $A$ is a $2 \times 3$ matrix and $B$ is a $3 \times 4$ matrix, then product $AB$ is a $2 \times 4$ matrix. To determine, for example the entry in row 2 and column 3 of $AB$, we single out row 2 from $A$ and column 3 from $B$. 
Example: Multiplying Matrices—Continued [1]

Then, as illustrated below, we multiply corresponding entries together and add up these products.

\[
\begin{pmatrix}
1 & 2 & 4 \\
2 & 6 & 0
\end{pmatrix}
\begin{pmatrix}
4 & 1 & 4 & 3 \\
0 & -1 & 3 & 1 \\
2 & 7 & 5 & 2
\end{pmatrix}
= \begin{pmatrix}
? & ? & 26 & ?
\end{pmatrix}
\]

The computations for the remaining entries are

\[
\begin{align*}
(1 \cdot 4) & + (2 \cdot 0) + (4 \cdot 2) = 12 \\
(1 \cdot 1) & - (2 \cdot 1) + (4 \cdot 7) = 27 \\
(1 \cdot 4) & + (2 \cdot 3) + (4 \cdot 5) = 30 \\
(2 \cdot 4) & + (6 \cdot 0) + (0 \cdot 2) = 8 \\
(2 \cdot 1) & + (6 \cdot 1) + (0 \cdot 7) = -4 \\
(2 \cdot 3) & + (6 \cdot 1) + (0 \cdot 2) = 12
\end{align*}
\]

and

\[
AB = \begin{pmatrix}
12 & 27 & 30 & 13 \\
8 & -4 & 26 & 12
\end{pmatrix}
\]
The System of Equations (1/14)

Consider one equation

\[ y(t) = a + bx(t) \]

\[ \Rightarrow y(t) = (1 \times x(t)) \begin{pmatrix} a \\ b \end{pmatrix} \]

\[ \Rightarrow y(t) = X(t)\beta \]

Then, we have the following equations

\[
\begin{cases}
\text{price}_A(t) \approx (1 \text{ factor}(t)) \begin{pmatrix} a_A \\ b_A \end{pmatrix} \\
\text{price}_B(t) \approx (1 \text{ factor}(t)) \begin{pmatrix} a_B \\ b_B \end{pmatrix}
\end{cases}
\]

The System of Equations (2/14)

**Definition: Linear Equation [1]**

We define a linear equation in the \( n \) variables \( x_1, x_2, \ldots, x_n \) to be one that can be expressed in the form

\[ a_1x_1 + a_2x_2 + \cdots + a_nx_n = b, \]

where \( a_1, a_2, \ldots, a_n \) and \( b \) are constants, and the \( a \)'s are not all zeros.

**Definition: System of Linear Equations [1]**

A finite set of linear equations is called a system of linear equations or, more briefly, a linear system.
Example: a General Linear System [1]

A general linear system of \( m \) equations in the \( n \) unknowns \( x_1, x_2, \ldots, x_n \) can be written as

\[
\begin{align*}
    a_{11}x_1 &+ a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\
    a_{21}x_1 &+ a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\
    \vdots & \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \ Quad
This is, we have the following equation

\[
\begin{align*}
\begin{pmatrix}
\text{price}_A(1) \\
\vdots \\
\text{price}_A(T) \\
\text{price}_B(1) \\
\vdots \\
\text{price}_B(T)
\end{pmatrix}
&= 
\begin{pmatrix}
1 & x(1) \\
\vdots & \vdots \\
1 & x(T)
\end{pmatrix}
\begin{pmatrix}
a_A \\
b_A \\
a_B \\
b_B
\end{pmatrix},
\end{align*}
\]

Consider two equations

\[
\begin{align*}
y_A(t) &= a_A + b_A x(t), \\
y_B(t) &= a_B + b_B x(t)
\end{align*}
\]

\[\implies \begin{pmatrix} y_A(t) \\ y_B(t) \end{pmatrix} = \begin{pmatrix} 1 & x(t) \end{pmatrix} \begin{pmatrix} a_A & a_B \\ b_A & b_B \end{pmatrix} \]

\[\implies y(t) = X(t)\beta.\]

That is, we have the following equation

\[
\begin{pmatrix}
\text{price}_A(t) \\
\text{price}_B(t)
\end{pmatrix}
\approx 
\begin{pmatrix} 1 & \text{factor}(t) \end{pmatrix}
\begin{pmatrix} a_A & a_B \\ b_A & b_B \end{pmatrix}.
\]
We can rewrite the above equation as

\[
\begin{pmatrix}
\text{price}_A(t) \\
\text{price}_B(t)
\end{pmatrix} \approx
\begin{pmatrix}
1 \text{ factor}(t) & 0 & 0 \\
0 & 0 & 1 \text{ factor}(t)
\end{pmatrix}
\begin{pmatrix}
a_A \\
\beta_A \\
a_B \\
\beta_B
\end{pmatrix}
\]

OK, this form is \( y = X\beta \).
Example: An Example for Kronecker Product [3]

The Kronecker product of $A = \begin{pmatrix} 3 & 4 & -1 \\ 2 & 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 5 & -1 \\ 3 & 3 \end{pmatrix}$ is

$$A \otimes B = \begin{pmatrix} 15 & -3 & 20 & -4 & -5 & 1 \\ 9 & 9 & 12 & 12 & -3 & -3 \\ 10 & -2 & 0 & 0 & 0 & 0 \\ 6 & 6 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Example: An Example for Kronecker Product–Continued [3]

The Kronecker product of $B = \begin{pmatrix} 5 & -1 \\ 3 & 3 \end{pmatrix}$ and $A = \begin{pmatrix} 3 & 4 & -1 \\ 2 & 0 & 0 \end{pmatrix}$ is

$$B \otimes A = \begin{pmatrix} 15 & 20 & -5 & -3 & -4 & 1 \\ 10 & 0 & 0 & -2 & 0 & 0 \\ 9 & 12 & -3 & 9 & 12 & -3 \\ 6 & 0 & 0 & 6 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 15 & 20 & -5 & -3 & -4 & 1 \\ 10 & 0 & 0 & -2 & 0 & 0 \\ 9 & 12 & -3 & 9 & 12 & -3 \\ 6 & 0 & 0 & 6 & 0 & 0 \end{pmatrix}.$$
**Definition: vec Operator [3]**

Let \( A = (a_1, \ldots, a_n) \) be an \((m \times n)\) matrix with \((m \times 1)\) columns \(a_i\). The **vec operator** transforms \( A \) into an \((mn \times 1)\) vector by stacking the columns, that is,

\[
\text{vec}(A) = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.
\]

**Example: An Example for vec Operator [3]**

For instance, if \( A = \begin{pmatrix} 3 & 4 & -1 \\ 2 & 0 & 0 \end{pmatrix} \) and \( B = \begin{pmatrix} 5 & -1 \\ 3 & 3 \end{pmatrix} \), then

\[
\text{vec}(A) = \begin{pmatrix} 3 \\ 2 \\ 4 \\ 0 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 4 \\ 0 \\ -1 \\ 0 \end{pmatrix} \quad \text{and} \quad \text{vec}(B) = \begin{pmatrix} 5 \\ 3 \\ -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \\ -1 \\ 3 \end{pmatrix}.
\]
Definition: Identity (or Unit) Matrix

A matrix $I = I_n$ is called identity (or unity) matrix if

$$
\begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{pmatrix}
$$

Rule: the $\otimes$ and vec Operator [3]

Let $A$, $B$, $C$ be matrices with appropriate dimensions. We have

$$
\text{vec}(AB) = (I \otimes A) \text{vec}(B).
$$

That is, we have the following equation

$$
\text{vec} \left((\text{price}_A(t) \ \text{price}_B(t))\right) \approx (I \otimes (1 \ \text{factor}(t))) \text{vec} \left(\begin{pmatrix} a_A & a_B \\ b_A & b_B \end{pmatrix}\right)
$$

and

$$
\text{vec} \left(\begin{pmatrix} \text{price}_A(1) & \text{price}_B(1) \\ \vdots & \vdots \\ \text{price}_A(T) & \text{price}_B(T) \end{pmatrix}\right) \approx (I \otimes (1 \ \text{factor}(1))) \cdots (I \otimes (1 \ \text{factor}(T))) \text{vec} \left(\begin{pmatrix} a_A & a_B \\ b_A & b_B \end{pmatrix}\right)
$$
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

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

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 the inverse of $A$. If no such matrix $B$ can be found, then $A$ is said to be singular.
Example: Invertible Matrix [1]

Let

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

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.

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.
Solution vs. Estimator (5/8)

Consider

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

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

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

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

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,

$$\begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_T \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \end{pmatrix} \quad \Rightarrow \quad y = X\beta + \epsilon$$

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

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

$$\Rightarrow \quad \beta = (X'X)^{-1}X'y - (X'X)^{-1}X'\epsilon$$
### Definition: Full Column Rank

Let \( A = (a_1, \ldots, a_n) \) be an \((m \times n)\) matrix with \((m \times 1)\) columns \(a_i\). Given an index \(i\), there exist \(b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_n\) such that

\[
a_i = b_1a_1 + \cdots + b_{i-1}a_{i-1} + b_{i+1}a_{i+1} + \cdots + b_na_n
\]

Then, the matrix \(A\) is not of full column rank.

This section is based on [2].

### Fundamental Assumptions

The ordinary linear regression model is described by the equation

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

where

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

The linear regression with fundamental assumptions is also called classical linear regression model.
Definition (3/4)

(a) Histogram Plot  (b) 2-D Line Plot

Figure: A Classical Linear Regression Model Example

Definition (4/4)

Figure: This X can include commodities prices, product prices, etc.
Notation: Unsolvale 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 $\beta_* \in \mathbb{R}^p$.

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.
Theorem: 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 $\alpha = g(\beta)$,

$$\frac{\partial h(g(\beta))}{\partial \beta'} = \frac{\partial h(\alpha)}{\partial \alpha'} \frac{\partial g(\beta)}{\partial \beta'}.$$
Proof.

By the differentiating the function \( f(\beta_*) \) with respect to \( \beta_* \),
\[
\frac{\partial f(\beta_*)}{\partial \beta_*} = \frac{\partial (y - X\beta_*)'}{\partial \beta_*} \frac{\partial (y - X\beta_*)'}{\partial (y - X\beta_*)} = (-X')(y - X\beta_*)'(I' + I) = -2X'y + 2X'X\beta_*.
\]

and \( \frac{\partial^2 f(\beta_*)}{\partial \beta_*^2} = 2X'X \) 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 \( \beta_* \).

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,

\[
\log L_T(\beta^*_*, \sigma^2) = -T \log \sigma \sqrt{2\pi} - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_t - X_t \beta^*_*)^2
\]

\[
= -T \log \sigma \sqrt{2\pi} - \frac{(y - X \beta^*_*)'(y - X \beta^*_*)}{2\sigma^2}
\]

\[
= -T \log \sigma \sqrt{2\pi} - \frac{\|y - X \beta^*_*\|^2}{2\sigma^2}.
\]

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

---

**Property (1/3)**

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

Under the linear regression model with fundamental assumptions,

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

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

**Proof.**

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

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

and

\[
\text{Cov} [\hat{\beta}] = E [\hat{\beta}\hat{\beta}^T - \beta\beta^T] = \sigma^2 (X'X)^{-1}.
\]
**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 $\text{plim}_{n \to \infty} \hat{\beta} = \beta$ 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 $\beta$. 

Kuan-Lun WANG  Fundamental Analysis of Securities Trading
