Comparison between Two-Stage Regression Model and Variance Model in Portfolio Optimization

Lam Weng Siew, Saiful Hafizah Hj. Jaaman and Hamizun bin Ismail

School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

**ARTICLE INFO**

- **Article history:** Received 25 July 2014
- **Received in revised form:** 8 July 2014
- **Accepted 15 September 2014**
- **Available online 17 October 2014**

**Keywords:** Index tracking Optimal portfolio Regression Variance Tracking Error

**ABSTRACT**

**Background:** Investors and fund managers aim to achieve similar return with the stock market index return in an investment. This can be done by purchasing all stocks in the market index with same proportion. However, this strategy is not practical since it incurs high transaction cost. Therefore, index tracking was introduced to construct an optimal portfolio which contains subset of stocks from the market index to achieve similar or approximate return with the market index. **Objective:** The objective of this paper is to construct an optimal portfolio using two-stage regression model and variance model in index tracking. The portfolio performance of both models are determined and compared in terms of tracking error. In this study, the data consists of 24 component stocks in Malaysia market index which is FTSE Bursa Malaysia Kuala Lumpur Composite Index from January 2010 until December 2012. **Results:** The result of this study indicates that both optimal portfolios constructed using variance model and two-stage regression model are able to track the market index effectively with only selecting 40% stocks from the market index. Besides, the optimal portfolio of variance model outperforms two-stage regression model due to lower tracking error. **Conclusion:** The optimal portfolio of variance model outperforms two-stage regression model in tracking Malaysia stock market index. Therefore, the variance model is more appropriate for the investors in Malaysia as compared to two-stage regression model.

© 2014 AENSI Publisher All rights reserved.

**To Cite This Article:** Lam Weng Siew, Saiful Hafizah Hj. Jaaman and Hamizun bin Ismail., Comparison between Two-Stage Regression Model and Variance Model in Portfolio Optimization. *J. Appl. Sci. & Agric.*, 9(18): 36-40, 2014

**INTRODUCTION**

Portfolio is a collection of stocks owned by the investors to generate return at minimum risk in stock market investment. A stock market index consists of the stocks with high capital which measures the performance of the major capital and industry segments of the country. One of the investment strategies is to track the stock market index in order to generate similar return with the index. The investors and fund managers can purchase all the stocks in the stock market index with same proportions. However, this strategy is not practical because it incurs high transaction cost to the investors. Therefore, Roll (1992) introduced index tracking in stock market investment. Index tracking is a portfolio management which aims to construct an optimal portfolio to generate similar return with the stock market index with only selecting some of the stocks from the market index. Roll (1992) presented a tracking error variance model in index tracking. Tracking error is a risk measure of how closely the portfolio follows the benchmark market index. Alexander (2005) proposed cointegration model in constructing the optimal portfolio in index tracking. The cointegration model has been studied by different researches in different stock market index. (Subramanian, 2008, and Grobys, 2011) Canakgoz and Beasley (2009) presented two-stage regression model in index tracking. This model is a mixed integer programming model which adopts regression approach. The objective of this paper is to construct an optimal portfolio using two-stage regression model and variance model in index tracking. The portfolio performance of both models are determined and compared in terms of tracking error. The rest of the paper is organized as follows. The next section describes the data and methodology. Section 3 discusses about the empirical results of this study. Section 4 concludes the paper.

**Data and Methodology:**

FTSE Bursa Malaysia Kuala Lumpur Composite Index (FBMKLCI) is the leading indicator of the performance of the Malaysia stock market and economy which consists of 30 stocks listed on the Malaysian

**Corresponding Author:** Lam Weng Siew, School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia 43600 UKM, Bangi, Selangor, Malaysia

E-mail: alexlam522@yahoo.com
Main Market. In this study, the data consists of weekly price of 24 stocks in FBMKLCI from January 2010 until December 2012. These 24 stocks are selected in the study since they make up as components of FBMKLCI consistently within the study period. In this study, the optimal portfolio is constructed using the variance model and two-stage regression model. The optimal portfolio is a group of stocks which satisfies the objective function and all the constraints in the mathematical model. The optimal portfolios of both models are computed using LINGO and EVIEWS software.

Tracking Error:
Roll (1992) applied variance as a risk measure for tracking error in index tracking. This is because variance is the most commonly used risk measure in portfolio optimization (Markowitz, 1952). Tracking error has been used as a risk measure in different index tracking models (Beasley et al., 2003, Wu et al., 2007, Lam et al., 2013). Tracking error is the standard deviation of the difference between the returns of the portfolio and the returns of the stock market index (Meade and Salkin, 1990). The formula for the tracking error is as follows.

$$TE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (R_{pt} - R_{it})^2}$$  

where $TE$ is the tracking error, $n$ is the number of periods, $R_{pt}$ is the mean return of the portfolio at time $t$ and $R_{it}$ is the mean return of the stock market index at time $t$. A portfolio is considered tracking the stock market index perfectly if there is zero tracking error. The optimal level of tracking error is below 3% (Fabozzi et al., 2002). The performance of the optimal portfolio constructed by the mathematical model is compared to the stock market index in terms of tracking error. In this study, the optimal portfolio is constructed using the variance model and two-stage regression model.

Variance Model:
Roll (1992) proposed the tracking error variance model in index tracking problem. The objective of the variance model is to minimize the tracking error subjects to five constraints. The model is formulated as below:

Minimize $TE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (R_{pt} - R_{it})^2}$  

Subject to

$$Z_i \in [0,1]$$  

$$\sum_{i=1}^{n} Z_i = K$$  

$$L_i Z_i \leq x_i \leq U_i Z_i$$  

$$0 < L_i < U_i < 1$$  

$$\sum_{i=1}^{n} x_i = 1$$

where $TE$ is the tracking error, $T$ is the number of periods, $R_{pt}$ is the mean return of the portfolio at time $t$ and $R_{it}$ is the mean return of the market index at time $t$. $x_i$ is the weight of each stock invested, $K$ is the number of stocks selected to track the market index. $L_i$ and $U_i$ are the lower and upper bounds of the investment proportion respectively on stock $i$.

Equation (2) is the objective function of the model which minimizes the tracking error. For constraint (3), $Z_i (i = 1,2,\ldots,n)$ is introduced to indicate the stock selection problem with $Z_i = 1$ indicates the $i$th stock is included in the tracking portfolio or otherwise for $Z_i = 0$. Constraint (4) ensures that the number of stocks in the tracking portfolio is $K$. In this study, $K$ is set to 12 which is 40% of the stocks in FBMKLCI index. (Jansen and Dijk, 2002). Constraint (5) shows that if stock $i$ is not selected in the tracking portfolio (i.e., $Z_i = 0$), then $x_i = 0$, and if stock $i$ is selected in the tracking portfolio (i.e., $Z_i = 1$), then $x_i \neq 0$. Constraint (6) indicates that the value of $x_i$ is limited in the interval $[L_i, U_i]$. Constraint (7) ensures that the total weight of stocks invested is one.

Two-Stage Regression Model:
Canakgoz and Beasley (2009) introduced mixed-integer programming model which adopts ordinary least square regression approach in index tracking. Regression is a statistical methods used to describe the nature of
the relationship between the variables. When a linear regression is performed on the return from the tracking portfolio against the return from the index, the ordinary least-square estimates, \( \hat{\alpha} \) and \( \hat{\beta} \), for the intercept and slope respectively of the regression line are given as below.

\[
\hat{\alpha} = \sum_{i=1}^{K} w_i \hat{\alpha}_i
\]

(8)

\[
\hat{\beta} = \sum_{i=1}^{K} w_i \hat{\beta}_i
\]

(9)

where \( w_i \) is the weight of stock \( i \) in the tracking portfolio, \( \hat{\alpha}_i \) and \( \hat{\beta}_i \) are the ordinary least-squares regression intercept and slope respectively when regression is performed on the returns from stock \( i \) against the index returns \( R_t \). In order to track the index perfectly, the target values for both estimates are \( \hat{\alpha} = 0 \) and \( \hat{\beta} = 1 \). There are two stages involved in the optimization process using two-stage regression model. The first stage aims to achieve the desired slope of one whereas the second stage aims to achieve the desired intercept of zero. Therefore, the objective of the model is to minimize \( |\beta - 1| \) and \( |\alpha - 0| \) in two stages. The model is formulated in two stages as shown below.

**First stage:**

Minimize \( |\beta - 1| \)

Subject to

\[
\sum_{i=1}^{K} z_i = K
\]

(10)

\[
\sum_{i=1}^{K} V_{iT} x_i = C
\]

(11)

\[
w_i = \frac{V_{iT} x_i}{C}
\]

(12)

\[
\sum_{i=1}^{K} w_i = 1
\]

(13)

\[L_i z_i \leq w_i \leq U_i z_i\]

(14)

\[
z_i \in [0,1]
\]

(15)

\[
x_i, w_i \geq 0
\]

(16)

**Second stage:**

Minimize \( |\alpha - 0| \)

\[
\sum_{i=1}^{K} z_i = K
\]

(17)

\[
\sum_{i=1}^{K} V_{iT} x_i = C
\]

(18)

\[
w_i = \frac{V_{iT} x_i}{C} = 1
\]

(19)

\[L_i z_i \leq w_i \leq U_i z_i\]

(20)

\[
z_i \in [0,1]
\]

(21)

\[
\hat{\beta} = \beta^{opt}
\]

(22)

\[
x_i, w_i \geq 0
\]

(23)

where \( K \) is number of stocks selected to track the stock market index, \( L_i \) and \( U_i \) are the lower and upper bounds of the investment proportion respectively on stock \( i \), \( V_{iT} \) is the price of one unit of stock \( i \) at time \( T \). \( x_i \)
is the number of units of stock $i$ in tracking portfolio, $C$ is the total amount invested at time $T$, $w_i$ is the weight of each stock invested and $\hat{\beta} = \beta_{mot}$ is the optimal value of the slope obtained in the first stage model.

Constraint (17) ensures that the number of stocks in the tracking portfolio equals to $K$. In this study, $K$ is set to 12 which is 40% of the components in the index (Jansen and Dijk, 2002). Constraint (18) defines the total amount invested at time $T$ for the tracking portfolio. Constraint (19) defines the weights of stock $i$ in tracking portfolio. For constraints (20) and (21), $z_i (i = 1, 2, ..., N)$ is introduced to indicate the stock selection problem with $z_i = 1$ indicates the $i$th stock is included in the tracking portfolio or otherwise for $z_i = 0$. Constraint (22) is the optimal value of the slope obtained in the first stage model. Constraint (23) ensures $x_i, w_i \geq 0$ are non-negative.

**Results:**

Table 1 presents the optimal portfolios compositions which are constructed using both variance model and two-stage regression model.

<table>
<thead>
<tr>
<th>Stock</th>
<th>Weights (%) (Variance Model)</th>
<th>Weights (%) (Two-Stage Regression Model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 AMMB Holdings</td>
<td>9.19</td>
<td>15.00</td>
</tr>
<tr>
<td>2 Axiata Group Bhd</td>
<td>11.42</td>
<td>15.00</td>
</tr>
<tr>
<td>3 British American Tabaco</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4 CIMB Group Holding</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5 Digi.Com</td>
<td>5.75</td>
<td>2.79</td>
</tr>
<tr>
<td>6 Genting</td>
<td>8.31</td>
<td>-</td>
</tr>
<tr>
<td>7 Genting Malaysia</td>
<td>3.67</td>
<td>-</td>
</tr>
<tr>
<td>8 Hong Leong Bank Bhd</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9 Hong Leong Financial Group</td>
<td>-</td>
<td>15.00</td>
</tr>
<tr>
<td>10 IOI Corporation</td>
<td>8.52</td>
<td>2.91</td>
</tr>
<tr>
<td>11 Kuala Lumpur Kepong</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12 Malayan Banking</td>
<td>11.15</td>
<td>15.00</td>
</tr>
<tr>
<td>13 Maxis Bhd</td>
<td>-</td>
<td>1.30</td>
</tr>
<tr>
<td>14 Petronas Dagangan Bhd</td>
<td>-</td>
<td>1.00</td>
</tr>
<tr>
<td>15 Petronas Gas Bhd</td>
<td>5.42</td>
<td>1.00</td>
</tr>
<tr>
<td>16 Public Bank</td>
<td>13.81</td>
<td>15.00</td>
</tr>
<tr>
<td>17 Public Bank Bhd</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18 RHB Capital Bhd</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>19 Sime Darby</td>
<td>12.46</td>
<td>15.00</td>
</tr>
<tr>
<td>20 Telekom Malaysia Bhd</td>
<td>-</td>
<td>1.00</td>
</tr>
<tr>
<td>21 Tenaga Nasional</td>
<td>5.79</td>
<td>-</td>
</tr>
<tr>
<td>22 UMW Holdings</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>23 YTL Corporation</td>
<td>4.52</td>
<td>-</td>
</tr>
<tr>
<td>24 YTL Power International</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Based on Table 1, there are only 12 stocks out of total sample 24 stocks selected in the optimal portfolio by both variance model and two-stage regression model. The optimal portfolio consists of 12 stocks with different weights to track FBMKLCI index which consists of 30 stocks. For the optimal portfolio of variance model, Public Bank (13.81%) is the most dominant stock whereas Genting Malaysia (3.67%) has the smallest component. For the optimal portfolio of two-stage regression model, AMMB Holdings, Axiata Group Berhad, Hong Leong Financial Group, Malayan Banking, Public Bank and Sime Darby (15.00%) are the most dominant stocks whereas Petronas Dagangan Berhad, Petronas Gas Berhad and Telekom Malaysia Berhad (1.00%) have the smallest components. Table 2 displays the comparison of performance between the optimal portfolio of variance model and two-stage regression model against FBMKLCI index.

<table>
<thead>
<tr>
<th>FBMKLCI Index (Benchmark)</th>
<th>Number of stocks</th>
<th>Tracking Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FBMKLCI</td>
<td>30</td>
<td>-</td>
</tr>
<tr>
<td>Variance Model</td>
<td>12</td>
<td>0.27</td>
</tr>
<tr>
<td>Two-Stage Regression Model</td>
<td>12</td>
<td>0.50</td>
</tr>
</tbody>
</table>

As shown in Table 2, the optimal portfolios of both variance model and two-stage regression model consist of 12 stocks to track FBMKLCI index which comprises 30 stocks. This implies that there is only 40% of FBMKLCI components are selected to construct the optimal portfolio. The tracking errors for variance model and two-stage regression model are 0.27% and 0.50% respectively which are closer to zero and well below
3.00%. This implies that the optimal portfolios constructed using both models are able to track FBMKLCI index effectively with only selecting 40% of the index components. Variance model outperforms two-stage regression model in index tracking due to lower tracking error.

**Conclusion:**
This paper discusses about the construction of the optimal portfolios using variance model and two-stage regression model to track FBMKLCI index. Both models are able to track the market index effectively with only selecting 40% of the index components. In index tracking, the investors are expected to generate similar return with the market index return without purchasing all the stocks in the market index using both models. Variance model outperforms two-stage regression model due to lower tracking error. In conclusion, variance model is more appropriate for the investors in Malaysia as compared to two-stage regression model.

**REFERENCES**