A panel data analysis of tourism and economic development in Southeast Asian countries

K Yunitaningtyas*, A M Yolanda, Indahwati

Department of Statistics, IPB University, Jln. Meranti Wing 22 Level 4 Kampus IPB Dramaga, Bogor 16680, Jawa Barat, Indonesia

Corresponding author: kristiana_yunitaningtyas@apps.ipb.ac.id

Abstract. Southeast Asia is a strategic region in tourism because of its natural and cultural richness. Thus, tourism ministers of ASEAN countries agreed to launch a joint ten-year plan to make Southeast Asia as one destination package for international tourists. The aim of this plan is to increase the share of tourism to the regional economy by 15 percent in 2025. This paper estimates the relationship between tourism and economic development in ten Southeast Asian countries of Brunei Darussalam, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, and Vietnam using regression analysis of panel data. It tests the effect of tourism factors such as visitor exports, internal travel and tourism consumption, business tourism spending and capital investment in tourism industry on gross domestic product from 2014 to 2016. The best model obtained is fixed effect model and the tourism aspects that have positive significant impact on gross domestic product improvement in Southeast Asian countries are internal travel and tourism consumption also capital investment in tourism industry. Furthermore, the factors that effect gross domestic product negatively are visitor exports and business tourism spending.

1. Introduction
Southeast Asia is one of the most strategic regions in the world because of its natural and cultural resources. Therefore in 2016 the tourism ministers from 10 countries in Southeast Asia submitted a joint plan for the next ten years to make Southeast Asia as one destination package for international tourists. The aim of this plan is to increase the share of tourism to the regional economy by 15 percent in 2025. Tourism is one of the main pillars of this region with 600 million inhabitants. Graph 1 illustrates the direct contribution of the travel and tourism sector to Gross Domestic Product (GDP). There has been an increase from year to year, especially in 2014 until 2017 and is predicted to rise continually to more than US $ 200 billion in 2027. The same thing was also shown in the percentage of contribution of the travel and tourism sector to GDP which exceeded 5 percent in 2014 to 2017.
According to the World Travel & Tourism Council’s (WTTC) data, the total contribution of Southeast Asian tourism sector to GDP is among the top five in the world at US$ 301.1 billion. This figure is the fastest growth among other regions. In addition, this sector was also able to produce 30,154,900 jobs until 2016. This proves that tourism is an important factor that increases the success of Southeast Asian countries in the economic field.

In this study four tourism aspects were observed on economic growth of Southeast Asian countries namely, Brunei Darussalam, Cambodia, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand and Vietnam from 2014 to 2016. These four aspects of tourism are expected to influence the GDP. The data is in the form of panels where the same cross-sectional unit is examined for a certain period of time or in short has a dimension of space and time[1]. This definition is in accordance with the data used because of the tourism aspect as well as GDP were observed for three consecutive years.

Previous studies have been conducted using panel data analysis especially to examine factors that influence different sectors among several regions or countries. For example Onaran[2] explores employment developments in Central and East European countries and the effects of integration with the world economy, Torres et. al[3] using panel estimation to shows the impact of oil abundance on economic growth, Torre and Myrskyla[4] tested the income inequality-health hypothesis using panel data over years 1975-2006 for 21 developed countries, and Narayan et. al [5] used panel data analysis to see the impact of tourism to the economic growth of Pacific Island countries.

The methods that can be used for analysis and estimation of panel data regression models are the common-effect, random-effects, and fixed-effect. The common-effect method combines time series and cross section data without looking at differences between time and individuals. In the fixed-effect method individuals effects are assumed to be fixed. While random-effects is a method where individual effects are assumed to be random variables that do not correlate with explanatory variables [6]. These three methods were applied in this study to get the best model that describes the influence of tourism to GDP in Southeast Asia.

**Figure 1.** Southeast Asia direct contribution of travel and tourism to GDP (source: Travel & Tourism Economic Impact 2017).
2. Panel Data Analysis

Panel data consists of a set of cross-sectional units which are observed from time to time and a combination of time series and cross-section data. In other words, panel data has observations on individual micro-units who are followed over time [7].

Panel data has several benefits such as it is controlling for individual heterogeneity which time series can not accommodate, panel data give more informative, more variability, less collinearity among the variables, more degrees of freedom and more efficiency [8]. Panel data allowed us to control for all shared period factors and time-invariant country-specific factors [4]. If each cross-sectional unit has the same number of time series observations, the panel data is called a balanced panel and if the number are different between observations it is called unbalanced panel.

Panel data regression model is different compared to time series regression and cross-section regression models, given by:

\[ Y_{it} = \beta_1 + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + \beta_5 X_{5it} + \epsilon_{it} \]  

Where \( Y \) stands for GDP, \( \beta \) for regression coefficient, \( X \) for the four Tourism aspects in ten Southeast Asian countries, namely visitor exports, internal tourism and travel consumption, business tourism spending, and capital investment in tourism industry. With \( i \) is the cross-sectional units, stands for \( i \)th of the ten countries, \( i = 1, 2, \ldots, 10 \) and \( t \) represents the time period (2014 to 2016). It is assumed that there are at most \( N \) cross-sectional or observation units and at most a number of \( T \) time periods [9].

2.1. Common-effect Model

This model is also called Pooled Least Square and it is the simplest approach. The assumption contained in this model is that there is no difference between the value of intercepts and slopes in the regression results. In other words the coefficient is constant to both individual and time. The method used in common effect model is Ordinary Least Square (OLS) and the model equation is written as;

\[ Y_{it} = \beta_1 + \sum_{k=2}^{K} \beta_k X_{kit} + \epsilon_{it} \]  

with \( i \) is \( i \)th of the ten countries \((i = 1, 2, \ldots, 10)\), \( t \) represents the time period, and \( X_{kit} \) is the \( k \)th explanatory variables, which are the four tourism aspects. \( \beta_1 \) stands for intercept and \( \beta_k \) is the slope coefficient or slope of the regression line.

2.2. Fixed-effect Model

In the fixed-effect model there are differences between individuals but the slope remains constant. The estimated coefficients of the fixed-effect models cannot be biased because of omitted time-invariant characteristics [8]. The fixed-effect model of this study is:

\[ Y_{it} = \beta_{1i} + \sum_{k=1}^{K} \beta_k X_{kit} + \epsilon_{it} \]  

\( i = 1, 2, \ldots, 10 \) (each country in Southeast Asia), \( t \) is the time period, \( X_{kit} \) is the \( k \)th explanatory variables, \( \beta_{1i} \) stands for intercept and \( \beta_k \) is the regression line slope.

There is addition technique in this method, called Least Square Dummy Variable (LSDV). Dummy variable is a method to isolate individual or time specific effect in a regression model. The fixed effect model with dummy variables of each country over time \( (D_{it}) \) where intercepts are different for each countries [9]:

\[ Y_{it} = \alpha_0 + \alpha_i D_{it} + \sum_{k=1}^{K} \beta_k X_{kit} + \epsilon_{it} \]
2.3. Random-effects Model

The problems caused by the fixed effect model and dummy variables are the loss of degree of freedom and prevents the original model being identified. This model is using Generalized Least Square (GLS) method to estimate the panel data regression. There are two assumptions in the random effect model [10], first the intersep and slope are different for each individu or country in this study,

\[ Y_{it} = \beta_{1i} + \sum_{k=1}^{K} \beta_{k}x_{kit} + u_{it} \]  

(5)

and the second, intersep and slope different for individu and time,

\[ Y_{it} = \beta_{1i} + \sum_{k=1}^{K} \beta_{kit}x_{kit} + u_{it} \]  

(6)

3. Data and Methods

Data is obtained from World Travel and Tourism Council’s (WTTC)[11] and International Financial Statistics database. The response variable in this study is Gross Domestic Product in each country. GDP is one method for calculating national income. The predictor variables consist of four tourism aspects that affect GDP of Brunei Darussalam, Cambodia, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand and Vietnam, explained as follows.

Visitor exports is spending within the country by international tourists for both business and leisure trips, including spending on transport, but excluding international spending on education. Internal travel and tourism consumption is total revenue generated within a country by industries that deal directly with tourists. Business tourism spending is spending on business travel within a country by residents and international visitors. Investment or capital investment is total investment by all industries that are directly involved in the tourism sector[11].

The methods carried out in this study are implementing three estimation methods for panel data regression, models based on common-effect, fixed-effect, and random-effects. Then determine the most appropriate estimation model by conducting several tests: Chow test, Hausmann test, and Langrange Multiplier test. Afterwards testing the regression parameters which are simultaneous and partial tests to see the effect of predictor variables overall and individually. Finally interpreting the best model that describes the relationship between predictor variables and the response variable. Software Eviews and R are used to estimate the regression analysis of panel data between GDP and tourism factors in ten countries in Southeast Asia.

4. Results

4.1. Panel data models

The three methods above are applied to the data and the estimated result for each model are obtained in the following tables.

| Table 1. Common-effect model estimates |
|----------------------------------------|
| Variable                  | Coefficient | t-statistic | Prob.  |
|---------------------------|-------------|-------------|--------|
| Constant                  | -62088.18   | -1.421930   | 0.1674 |
| Visitor Exports           | -14.09302   | -1.166949   | 0.2542 |
| Investment                | 15.59604    | 1.605656    | 0.1209 |
| Internal T & T Consumption| 8.192611    | 1.022323    | 0.3164 |
| Business Spending         | 79.23592    | 7.849515    | 0.0000 |
Table 2. Fixed-effect model estimates

| Variable                             | Coefficient | t-statistic | Prob. |
|--------------------------------------|-------------|-------------|-------|
| Constant                             | -7783862    | -4.292112   | 0.0006|
| Visitor Exports                      | -41.27889   | -2.233890   | 0.0401|
| Investment                           | 285.0082    | 4.155272    | 0.0007|
| Internal T & T Consumption           | 96.42303    | 3.235229    | 0.0052|
| Business Spending                    | -293.3044   | -2.375440   | 0.0304|

Table 3. Random-effects model estimates

| Variable                             | Coefficient | t-statistic | Prob. |
|--------------------------------------|-------------|-------------|-------|
| Constant                             | -62088.18   | -2.726676   | 0.0115|
| Visitor Exports                      | -14.09302   | -2.237726   | 0.0344|
| Investment                           | 15.59604    | 3.078985    | 0.0050|
| Internal T & T Consumption           | 8.192611    | 1.960393    | 0.0612|
| Business Spending                    | 79.24592    | 15.05213    | 0.0000|

In panel data regression analysis, the models obtained can be derived from several methods thus the further testing is needed in choosing the right model to predict the regression.

4.2. Chow test

Chow test is conducted to test the fixed-effect or the joint significance of the dummies [12] with hypothesis:

\[ H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0 \]
\[ H_1 : \text{At least one } \alpha_i \text{ is different} \]

Chow test statistic:

\[ F = \frac{(RSSS - URSS)/(N - 1)}{URSS/(NT - N - K)} \]  

This F-test selects between common-effect and fixed-effect models with the restricted residual sum of squares (RRSS) being that of OLS on the pooled or common effect model and the unrestricted residual sum of squares (URSS) being that of the LSDV or fixed effect regression model [12]. \( H_0 \) is rejected if \( F \) less than \( F \) table or p-value is less than significance level (\( \alpha \)). When the Chow test shows that the corresponding model is fixed-effect, the next step is to conduct the Hausmann test. But when it shows the otherwise, then the significance test of the model are carried out [9].

With the help of Eviews program p-value obtained is 0.0000. Therefore the p-value is less than \( \alpha \) (0.05) so there is enough evidence to reject null hypothesis. It is undeniable that there is influence of individual effect on GDP in Southeast Asia region. Clearly it can be seen by looking at the heterogeneny between sectors in tourism and GDP. The appropriate model is the fixed-effect.

4.3. Hausmann test

The Hausmann test is used to compare the random-effects model to the fixed-effect models. The hypothesis for this test:

\[ H_0 : \text{correlation } (X_{ij}, u_{ij}) = 0 \text{ or model fits fixed-effect} \]
\[ H_1 : \text{correlation } (X_{ij}, u_{ij}) \neq 0 \text{ or model fits random-effects} \]

The statistic of Hausmann test:
\[ W = X^2(K) = (b - \beta)[\text{var}(b) - \text{var}(\beta)]^{-1}(b - \beta) \] (8)

\( H_0 \) is rejected when \( W \) is less than \( \chi^2 \) or p-value is less than significance level (\( \alpha \)) which means fixed-effect is the more appropriate model. The result of p-value obtained from the Eviews program is 0.0000, which is less than \( \alpha \) (0.05). This means \( H_0 \) is rejected and fixed-effect model is better than random-effect model in this study.

### 4.4. Lagrange Multiplier (Breusch-Pagan) test

The aim of this test is to see panel heteroscedasticity in the fixed-effect model. The hypothesis is given by:

\[ H_0 : \sigma_i^2 = 0 \text{ or no residuals heteroscedasticity} \]
\[ H_1 : \sigma_i^2 \neq 0 \text{ or there is residuals heteroscedasticity} \]

Lagrange Multiplier test statistic formula is written as:

\[ LM = \frac{KT}{2(T-1)} \left( \frac{\sum_{i=1}^{K} \sum_{t=1}^{T} e_{it}^2}{\sum_{i=1}^{K} \sum_{t=1}^{T} e_{it}^2} - 1 \right)^2 \sim \chi^2_{K-1} \] (9)

\( K \) is the number of sectors, \( T \) is the time period, and \( e_{it} \) is the residual of common-effect model.

When \( LM \) statistic value is more than \( \chi^2_{K-1} \) or a probability value (p-value) is less than \( \alpha \) the best model obtained is random effect [13].

| Table 4. Lagrange Multiplier test result. |
|------------------------------------------|
| Effect       | t-statistics | Probability |
|--------------|--------------|-------------|
| cross-section | 0.247386     | 0.6189      |
| time         | 0.500360     | 0.4793      |
| both         | 0.747745     | 0.3872      |

Based on the table shown above, probability values for cross-section is more than the significance level \( \alpha \) (0.05) so there is not enough evidence to reject \( H_0 \). Hence there is no heteroscedasticity in residuals.

### 4.5. Parameter regression tests

Testing the regression parameters needs to be done to determine the relationship between the response and predictor variables. Tests for regression parameters are carried out in two stages, namely joint test (simultaneous) and partial test.

Simultaneous tests are conducted to see whether predictor variables jointly influence the response variable. In this case it tests the overall aspects of tourism whether have an effect on the Southeast Asian countries GDP.

\[ H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \]
\[ H_1 : \text{At least one} \beta_k \neq 0, k = 1, 2, \ldots, K \]

Test statistic formula for simultaneous test given:

\[ F= \frac{\text{Mean Squares Regression}}{\text{Mean Squares Residuals}} \] (10)

Null hypothesis is rejected when \( F \) is less than \( F_{\text{table}}(F_{\alpha; (K-1,N-k)}) \) or p-value is less than significance level. The Eviews program output gives result that p-value is 0.0000. It means that there is enough evidence to reject null hypothesis and it can be concluded that the predictor variables together influence the response variable. In other words, the four aspects of tourism as a whole have an effect on total GDP.
The partial test is used to determine the predictor variables that have individually significant effect on the response variable, with the hypothesis:

\[ H_0 : \beta_k = 0 \]
\[ H_1 : \beta_k \neq 0, k = 1, 2, ..., K. \]

Test statistic formula for partial test given:

\[ t = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \]  \hspace{1cm} (11)

Null hypothesis is rejected when \(|t|\) is less than \(t_{table}(\alpha; N - K)\), with \(N\) stands for the number of observations and \(K\) is the number of parameters.

| Variable                | \(t\)-statistics | Prob.  |
|-------------------------|-------------------|--------|
| Visitor Exports         | -2.233890         | 0.0401 |
| Investment              | 4.155272          | 0.0007 |
| Internal T & T Consumption | 3.235229       | 0.0052 |
| Business Spending       | -2.375440         | 0.0304 |

The table shows that four tourism factors have probability values less than significance level \(\alpha\) (0.05). Thus visitor exports, investment, internal travel and tourism consumption, and business spending individually affect GDP.

5. Conclusion

The best model obtained in panel data regression analysis to predict the relationship between the tourism sector and the regional economic growth (GDP) of countries in the Southeast Asia region is the fixed-effect model with the equation:

\[ GDP_{it} = -7783862 + a_t - 41.27889 \ \text{Visitor Exports} + 285.0082 \ \text{Investment} + 96.42303 \ \text{Internal Travel and Tourism Consumption} - 293.3044 \ \text{Business Spending}. \]

The GDP of ten countries in Southeast Asia is directly proportional to the amount of investment by all industries directly involved in the tourism sector. Furthermore GDP also has positive relationship with the total income generated by all tourism industries. GDP is inversely proportional to the amount of domestic expenditure by foreign tourists and expenditure on travel domestic business by local residents and foreign visitors.

From the model, visitor exports are negatively associated with the GDP value of Southeast Asian countries. This is sensible because basically according to the UNWTO Recommendations on Tourism Statistics [14], visitor exports can be divided into three groups, particularly before, during and after the trip. Expenditure for preparation and after trip are made in the country of origin when preparing for and returning from a trip abroad. Business spending does not directly contribute to a country’s financial income as a result of people leaving immediately or moving to another place once their business is finished.

Therefore to increase the regional economic growth of each country in Southeast Asia as well as the overall economic growth, support from the government is needed in advancing the tourism industry, both through improving facilities and access and attracting investors to contribute more to the tourism sector. In the other hand, promotion of products and services produced by the local tourism industry also must be intensified. It can be done through advertising both inside and outside the country to increase the revenue of the industry. In addition, a joint plan between countries needs to be realized as an effort in supporting economic progress of all nations.
References
[1] Damodar G N 2004 Basic Econometrics (New York : McGraw-Hill)
[2] Onaran O 2008 Jobless growth in the Central and East European countries: A country-specific panel data analysis of the manufacturing industry East. Europ. Econ.46 90–115
[3] Torres N, Añonso Ó and Soares I 2012 Oil abundance and economic growth-a panel data analysis Energy J.33 119–48
[4] Torre R and Myrskylä M 2011 Income inequality and population health : a panel data analysis on 21 developed countries Demogr. Res.49 116
[5] Narayan P K, Narayan S, Prasad A and Prasad B C 2010 Tourism and economic growth : a panel data analysis for Pacific Island countries Tour. Econ.16 169–83
[6] Greene W 2000 Econometric analysis. 4th ed (New Jersey: Pearson Education, Inc.)
[7] Hill R C, Griffiths W E and Lim G C 2011 Principles of Econometrics (New Jersey : John Wiley & Sons, Inc.)
[8] Hsiao C 2004 Analysis of Panel Data (Cambridge University Press)
[9] Nwakuya M T and Ijomah M A 2017 Fixed effect versus random effects modeling in a panel data analysis; aconsideration of economic and political indicators in six African countries Int. J. Stat. Appl.7 275–9
[10] Lestari A and Setyawan Y 2017 Analisis regresi data panel untuk mengetahui faktor yang mempengaruhi belanja daerah di provinsi Jawa Tengah J. Stat. Ind. dan Komputasi2 1–11
[11] [World Travel and Tourism Council] 2017 Travel andtourism economic impact South East Asia
[12] Baltagi B H 2005 Econometric Analysis of Panel Data (West Sussex: John Wiley & Sons, Inc.)
[13] Rahmadeni and Yonesta E 2016 Analisis regresi data panel pada pemodelan produksi panen kelapa sawit di kebun sawit plasma Kampung Buatan Baru J. Sains Mat. dan Stat.2 1–12
[14] Weiß P 2006 Data sources on tourism expenditure. the austrian experiences taking into account the thop requirements. Int. Work. Tour. Stat.