Modelling and Forecasting Tourist Arrivals to Cambodia: An Application of ARIMA-GARCH Approach

Theara Chhorn and Chukiat Chaiboonsri

Faculty of Economics, Chiang Mai University

13 December 2017

Online at https://mpra.ub.uni-muenchen.de/83942/
MPRA Paper No. 83942, posted 17 January 2018 14:27 UTC
Modelling and Forecasting Tourist Arrivals to Cambodia: An Application of ARIMA-GARCH Approach

Theara Chhorn* and Chukiat Chaiboonsri**
*Faculty of Economics, Chiang Mai University
**Faculty of Economics, Chiang Mai University

Corresponding author: theara.chhorn.research@gmail.com

Please cite this article as:

ABSTRACT

The paper aims at estimating and forecasting international tourist arrivals to Cambodia during the time interval of 2000m1 to 2017m7, covering 209 of monthly observations. To find out factors affecting tourist arrivals, simple OLS and 2SLS with instrument variable regression are applied, on the one hand. On the other hand, several time series models of ARIMA (p, d, q), GARCH (s, r) and the hybrid of ARIMA(p, d, q)-GARCH(s, r) are employed to forecast tourist arrivals in line with AIC and BIC in selecting the best modified models. The empirical results primarily reveal that tourist arrivals are affected by exogenous factor, say exchange rate, dummy factors such as the AEC, global financial crisis, national election and Cambodia’s e-Visa. With regard to forecasting stage, the result indicates that tourist arrivals are shocked by time trend in the past period, say time (t-1). The trend is furthermore reduced due to the time lags, say time (t-2, t-3) as shown in the parameter coefficients of AR. GARCH (1, 1) model suggests that the short run persistence of shocks lies in the gap of 0.04 whereas the long run persistence lies in the gap of 0.94. Additionally, AIC and BIC propose that ARIMA(3, 1, 4) and the hybrid of ARIMA(3, 1, 4)-GARCH (1, 1) are the best model to predict the future value of tourist arrivals. The RMSE and U index obtained from measurement predictive accuracy reveal that long run 1-step ahead forecasting of 2013m12 to 2017m7 is produced the smallest predictive error amongst the others. Thus, it has more predictive power to apply long term ex-ante forecasting.

Key words: Point Forecasting Interval, out of Sample Forecasting, ARIMA (p, d, q)-GARCH (s, r) Model, Exchange rate and Dummy Factors, Tourist Arrivals, Cambodia

JEL Classification: C22, C53, Z3

TECHNICAL OBSERVATIONS

Prediction or forecasting is generally considered as an art of anticipation or estimation any future event and/or value. In the context of economic and financial time series analysis, it somehow takes into account the prediction methods due to econometric
models in line with statistical inferences. It is knowingly separated by two main categories that so-called in sample and out of sample forecasting or say ex-post and ex-ante forecasting. Good performance in out-of-sample prediction is viewed as the acid test for a good forecast model (Kunst 2012). It reflected the facts of any econometric models which is perfectly and methodologically adopted. In this case, diagnostic tests are employed conventionally. Forecasting in macroeconomic or financial data is widely acknowledged since it has played an important catalyst for policy makers as well as financial trader to set up the policy in achieving growth, development of the country and to gain profit from market speculation respectively. Prediction in general vastly meet the maturity. Many different approaches, both linearity and nonlinearity, due to the combination method of mathematics and/or statistical inferences are applied, (Wang 2016), (Sjo 2011), (Chia-Lin Chang 2009) and (Elzbieta F. and Mirosław Ga 2004). Time series models can be by definition giving a reasonable benchmark to evaluate the ranging value of forecasting based on periodical step ahead and/or full and/or sub sample observation relevant to the pure explanatory power of historical behavior of the series if the methodological assumption is detected and not violent the assumptions within the models such as the Box-Jenkins methodology of ARIMA (p, d, q), (Box and Pierce 1977).

In financial and economic time series estimation and prediction, the most common models which were typically and frequently employed are autoregressive conditional heteroskedasticity or so-called the ARCH model, (Baum 2015). Using an ARMA processes with up to two lags and variance with one of GARCH, EGARCH or TARCH processes with up to two lags, (Jánský 2011) evaluated several hundred one-day-ahead of VaR forecasting models. GARCH process as the best conditional volatility process for the analyzed time series, stated by the above author. It helped improvements, including a no prior assumption on the distribution of the log returns, which proved to be a step in the right direction. Consequently, time-variation in volatility (heteroskedasticity) is a common feature of financial data. The most straightforward way to measure the heteroskedasticity is to estimate the time-series of variances on “rolling samples”, (Chen 2013). This can model by considering heteroskedasticity. Yet, in the context of tourist arrival forecasting, it can be defined that as time-varying conditional variance has both the AR and MA components, it leads to the generalized ARCH (p,q) (GARCH (p,q)) of Bollerslev (1986), (Chia-Lin Chang 2009). In order to manage international tourism growth, therefore it is essential to model sufficiently tourist arrivals and their association due to volatility and autoregressive model and forecasting in the study of (Chia-Lin Chang 2009). Mamula (2015) suggested that although the diagnostics for the selected models reveals that the proposed models do not significantly differ, it can be concluded that the multiple regression model performs a highly accurate forecasting of tourist arrivals.

Tourism demand forecasting currently and widely employed the variety of forecasting methods, running from the most rudimentary approaches to the more complex, (Ramos 2014). Tourist arrivals and expenditure (receipt), in both aggregate and per capita forms, are commonly used to measure tourism demand in empirical research, (Song et al. 2010). In line to its modelling and forecasting, most of the empirical studies have focused on conventional approaches of forecasting performance toward measurement predictive accuracy, those models are included but not limited univariate ARMA and ARIMA based models, GARCH or the hybrid ARMA-GARCH, ARIMA-GARCH, seasonality components as well as nonlinear models, (Chia-Lin Chang 2009), (K.-Y.

With respect to few exogenous and dummy factors, (George Agiomirgianakis 2014), found the negative relationship between exchange rate volatility and tourist inflows into Turkey. The study of exchange rate volatility and tourism demand, (Webber 2001), (Yap 2012) and (Chang 2009). In contrast, (Crouch Geoffrey I. 1993) underlined that exchange rate volatility is a contributing factor to tourist arrivals. Both the moving average and the high and low measures of volatility have proven to have a significant effect to tourist arrivals. In addition, the impact of financial crises on tourism demand is less significant. Ensuring the safety and health of tourists is the key to maintain demand for inbound tourism, (Yu-Shan Wang 2009). Since the 2008 global financial crisis and resulting recession, many countries have been following unconventional monetary policies. Some findings highlight the importance of government economic policy in stimulating international tourism demand through its impact on the economy, (Jewoo Kim 2016).

The central purposes of the study are firstly to find out the factors which might affect to tourist arrivals in Cambodia. It aims secondly at modeling and forecasting tourist arrivals through the autoregressive and volatility approach. Within this stage, the study accordingly adopts 1-step ahead of out of sample forecasting to check the measurement predictive accuracy as well. Indeed, the remainder of the paper is organized as follows. The 1st session is to present the facts of time series prediction and factors affecting to tourist arrivals whereas the 2nd one is to design the methodology and data calculation. The 3rd and last session is to interpret the empirical outcomes and concluding remarks along with suggestions.

**METHODOLOGY AND DATA CALCULATION**

*Data Collection and Calculation*

Responding to our central objectives, international tourist arrivals in Cambodia, simply noted as (TA) is employed. Monthly observations of TA from 2000m1 to 2017m7 are extracted from ministry of tourism, (MoT, 2017). Exchange rate (EX) is imported from the CEIC data manager. The study transferred the sample observations of TA into the nature of log return using the formula as follows: \( y_t = \ln(TA_t) - \ln(TA_{t-1}) \) for the forecasting stage. Therefore, the descriptive statistics of both TA and EX with and without taking logarithm function are demonstrated in table 1 as bellows:

Table 1 shows the different types of descriptive statistics such as mean, minimum and maximum as well as LM test so on. With the total observations of 209, Shapiro-walk statistics\(^1\) and normality test show that the series TA and EX are not come from normal distribution assumption. LM test for ARCH effect somehow reveals the rejection of the null hypothesis at 1 % level of significance. It statistically defines that the series are contained ARCH effect due to the null hypothesis of no ARCH effect\(^2\). Skewness and

\(^1\) Shapiro - Wilk and Shapiro - Francia tests for normality, https://www.stata.com/manuals13/rswik.pdf

Kurtosis statistics give insights into the shape of the normal population distribution. Skewness essentially measures the relative size of the two tails or say the normal distribution has a Skewness of 0 whereas Kurtosis is a measure of the combined size of the two tails. Its value often compares to kurtosis which is equal to 3 or greater. As the result, both statistics indicate that Skewness value is ranged in 0.64 and -0.39 and -0.51 and -0.59 respectively whereas Kurtosis is 2.5 and 2.06 for nature and logarithm data of TA, respectively. It statistically indicates that the series, TA_t and ln(TA_t) is non-normal distribution as similar as denoted in the Shapiro-walk test as well.

Table 1: Descriptive statistics of tourist arrivals to Cambodia, TA

<table>
<thead>
<tr>
<th>Description</th>
<th>TA_t</th>
<th>ln(TA_t)</th>
<th>EX_t</th>
<th>ln(EX_t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>209</td>
<td>209</td>
<td>209</td>
<td>209</td>
</tr>
<tr>
<td>Percentiles (50%)</td>
<td>173112</td>
<td>12.0617</td>
<td>4104.482</td>
<td>8.3198</td>
</tr>
<tr>
<td>Mean</td>
<td>204768.4</td>
<td>11.9678</td>
<td>4089.954</td>
<td>8.3159</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>135157.1</td>
<td>7.782</td>
<td>115.868</td>
<td>0.0285</td>
</tr>
<tr>
<td>Min</td>
<td>30485</td>
<td>10.3249</td>
<td>3811.999</td>
<td>8.246</td>
</tr>
<tr>
<td>Max</td>
<td>611534</td>
<td>13.324</td>
<td>4305.74</td>
<td>8.368</td>
</tr>
<tr>
<td>Variance</td>
<td>1.83e+10</td>
<td>0.6056</td>
<td>13425.46</td>
<td>0.0008</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.6354</td>
<td>-0.3937</td>
<td>-0.5136</td>
<td>-0.588</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.5012</td>
<td>2.0681</td>
<td>3.004</td>
<td>3.088</td>
</tr>
<tr>
<td>Shapiro – Walk Test</td>
<td>5.4***</td>
<td>4.65***</td>
<td>4.22***</td>
<td>4.52***</td>
</tr>
<tr>
<td>LM test for ARCH effect</td>
<td>140.01***</td>
<td>174.29***</td>
<td>181.79***</td>
<td>183.19***</td>
</tr>
</tbody>
</table>

Source: Author’s estimates

Note: The sign notification of *, ** and *** referred to the statistical significance of 10%, 5% and 1%. LM test for ARCH considered 1 degree of freedom in order to test after a regression of its own trend. The value inside the parenthesis is the p-value.

Estimation and Prediction Approach

Baseline Regression Model of Tourist Arrivals in Cambodia

To estimate the factors affecting tourist arrivals in Cambodia during the period observations, let’s consider a sample regression equation in line with time trend effect, (TT_t) as follows:

\[ TA_t = c + \gamma[X'_t] + \tau[TT_t] + u_t, \sigma^2_t \sim N(0, 1) \] (1)

Where \( TA_t \) is an explained variable and it is denoted as international tourist arrivals. \( u_t \) is an error term and the constant term, (c). \( X'_t \) is a matrix set of explanatory variables. The study employs exchange rate as the main explanatory variable. It is known partially as tourism price effect. \( TT_t \) is a matrix set of time trend effect during the observed periods. To overcome the time trend effect on the model, dummy (binary option) variables which take into account number 1 for the determined period and 0 otherwise, will be adopted. Those dummy variables are included the national election in 2003, 2008 and 2013, the global financial crisis in 2008 and 2009, milestone of the ASEAN Economic Community (AEC) in 2015 and Cambodia first e-Visa launching from 2006 to present. These variables believe to have the strong impact to travel decision. Yet, the study applies OLS estimator with robust standard error (SE) and 2SLS with instrument variable (IV) to estimate the equation (1).

Furthermore, to model and forecast the volatility and the time trend effect of tourist arrivals individually, the study applies ARIMA (p, d, q), GARCH (s, r) and the hybrid
ARIMA (p, d, q)-GARCH (s, r) model. Therefore, ARIMA (p, d, q) is modeled as follows:

**ARIMA and ARIMA-GARCH Model**

Most of time series data are econometrically affected by either autoregressive process (AR) or moving average process (MA). Let’s consider, in one part, an ARIMA (p, d, q) with parameter order of p and q in line with d (order of integrated, D) as follows:

\[ y_t = \alpha_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varphi_1 u_{t-1} + \varphi_2 u_{t-2} + \cdots + \varphi_q u_{t-q} \]  

It is noteworthy that equation (2) can be derived with different data and lag operators or AR (p) and MA (q) process. Hence, we get:

\[ y_t = \alpha_t + \sum_{p=1}^{t} \phi_p y_{t-p} + \sum_{q=1}^{t} \varphi_q u_{t-q} + \tau T_t \]  

Or using the backshift notation with \( By_t = y_{t-1} \), the above equation, (3) can write as follows:

\[ (1 - \phi_1 B - \cdots - \phi_p B^p) y_t = (1 - \varphi_1 B - \cdots - \varphi_q B^q) u_t \]  

Furthermore, using the first different of series, \( W_t = y_t - y_{t-1} \) or \( W_t = (1 - B) y_t \), the specific general form of ARIMA (p, d, q) is equated as follows:

\[ \phi_p(B)(1 - B)^d y_t = \varphi_q(B) u_t \]  

Determining the parameter in different equation is a must in ARIMA (p, d, q) due to Box-Jenkins methodology, (Box 1970) and (Box and Pierce 1977). Accordingly, checking stationary or unit roots is essential and important to decide the parameters of all elements in the model. The study employs two popular methods of unit roots test, namely an augmented Dickey–Fuller (ADF) tests, (Dickey 1981) and Phillips-Perron (PP) test, (Phillips 1988). The study tests these two tests with and without trend and intercept. It is noted that these tests contain the null hypothesis of having unit root or meaning that the series is non-stationary. As the result, the empirical outcomes demonstrate in table 2.

**Table 2: Unit roots test analysis of TA and EX**

<table>
<thead>
<tr>
<th>Description</th>
<th>TA_t</th>
<th>ln(TA_t)</th>
<th>EX_t</th>
<th>ln(EX_t)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>At level, I(0)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADF</td>
<td>-1.76</td>
<td>-5.32</td>
<td>-1.82</td>
<td>-4.88</td>
</tr>
<tr>
<td></td>
<td>(0.4003)</td>
<td>(0.0001)</td>
<td>(0.3722)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>PP (Z(rho))</td>
<td>-6.28</td>
<td>-59.86</td>
<td>-5.01</td>
<td>-53.24</td>
</tr>
<tr>
<td></td>
<td>(0.3900)</td>
<td>(0.0000)</td>
<td>(0.3763)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td><strong>At first difference, I(1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADF</td>
<td>-12.5</td>
<td>-12.5</td>
<td>-11.98</td>
<td>-11.96</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>PP (Z(rho))</td>
<td>-160.48</td>
<td>-160.4</td>
<td>-146.06</td>
<td>-146.05</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
</tbody>
</table>

**Source:** Author’s estimates  
**Note:** The statistical value in the parenthesis is p-value. P-value of 0.0000 indicated the statistical significance of rejection the null hypothesis at 1%. NT is referred to the estimation without trend whereas IT with trend and intercept.

Without trend and intercept, both series are non-stationary at level and they are stationary at first different, say I(1) due to the converting. Conversely, with trend and intercept, the series is stationary at level, say I(0). As the result, to control the
interaction of log likelihood not to be concave iteration in post estimation, logarithm series with first difference will be employed. Next, the study presents a brief description of GARCH (s, r) model.

With regard to volatility approach, ARCH model introduced by (Engle 1982) and generalized ARCH, the so-called GARCH (Generalized ARCH) by (Bollerslev 1986) is used to investigate the volatility effect in the series, both low and high frequency data. The models widely adopt in various branches of econometrics, particularly in financial time series analysis. To estimate and forecast TA, a standard GARCH (1, 1) with no regressors in the mean and variance equations is proposed. Therefore, the model is equated as follows:

\[ Y_t = \theta X' t + \varepsilon_t, \varepsilon_t = \sigma_t z_t \]

\[ \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \]

Since \( \sigma_t^2 \) is the one-period ahead forecast variance based on past information, it is called the conditional variance. \( \varepsilon_{t-1}^2 \) and \( \sigma_{t-1}^2 \) are an ARCH and GARCH term respectively. Parameter testing states that \( z_t \) is standardized residual returns (i.e. iid random variable with zero mean and variance. For GARCH (1, 1), the constraints \( \alpha, \beta \geq 0 \) and \( \omega > 0 \) is needed to ensure that \( \sigma_t^2 \) is strictly positive, (Suliman 2011). As the result, from equation (5) and (7) we can derive the hybrid of ARIMA (p, d, q)-GARCH (1, 1) model as follows:

\[ \phi_p(B)(1 - B)^d y_t = \varphi_q(B)u_t + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \]

**Measurement Predictive Accuracy**

Since the paper aim at forecasting tourist arrival from post estimation and out of sample prediction, the measurement predictive accuracy is adopted. Hereafter, let’s assume forecast sample is \( j = T + 1, T + 2, ..., T + h \), and denote the actual and forecast value in period \( t \) as \( y_t \) and \( \hat{y}_t \), respectively. The study uses two types of error predictive method, namely root mean square error (RMSE) and Theil’s inequality index (U) to measure. Furthermore, it is used to specify the best model for the purpose of long run ex-ante prediction. Therefore, the forecast evaluation measurements, RMSE and U define as follows:

\[ \text{RMSE} = \sqrt{\frac{\sum_{t=T+1}^{T+h}(\hat{y}_t - y_t)^2}{h}} \]

\[ \text{Theil’s U} = \frac{\sqrt{\sum_{t=T+1}^{T+h}(\hat{y}_t - y_t)^2/h}}{\sqrt{\sum_{t=T+1}^{T+h}(\hat{y}_t)^2/h} + \sqrt{\sum_{t=T+1}^{T+h}(y_t)^2/h}} \]

Shortly, the study employs ARIMA (p, d, q), GARCH(s, r) and the hybrid of ARIMA (p, d, q)-GARCH (s, r) to model and forecast tourist arrivals in Cambodia from monthly observation of 2000m1 to 2007m7. The RMSE and U are employed to measure the predictive accuracy from out of sample forecasting.

**EMPIRICAL RESULT AND DISCUSSING**

This session aims at interpreting the empirical outcomes from post-estimation of the baseline regression equation, equation (1) and forecasting models from ARIMA (p, d, q), GARCH (s, r) and the hybrid of ARIMA (p , d, q)-GARCH (s, r). Indeed, the study
presents some calibration of tourism trend and development since 1993 to present toward numerical sources and graphical illusions.

Calibration of Cambodia’s Tourism Industry

Table 3 shows the tourism trend in Cambodia since 2013 till February 2017 and its growth rate year-on-year. From 2016 to the first quarter of 2017 (Q1), tourism growth rate approximates 12.1% comparing to those of 2015 and 2016 which presents 5% and 6.1% for 2015/2014. It demonstrates from year to year that tourism trend is increased considerably due to on the one hand government considers tourism industry as one of the central sector in contributing to growth and development. On the other hand, Tourism Development Strategic Plan 2012-2020 is implemented with the goal of attracting tourist throughout connectivity, safety and security, marketing and facilitation of tourist transportation, (MoT 2012) for example. In 2013, international tourist arrivals account over 4 million people while it reached up to almost 5 million people in 2016. It reflected the gap of slowing growth, as it presents just 1 million people coming to visit Cambodia for almost 4 years. According to the research, only 80% have visited Cambodia one time and 20% is more than one. Furthermore, lack of infrastructure, especially the sewage facilities, inadequate accommodations and facilities, personal security issues, and accessibility to secondary destinations are items which need immediate attention, (Paul Leung 2017).

Table 3: International tourist arrivals to Cambodia, 2003 - 2017

<table>
<thead>
<tr>
<th>Months</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
<th>15/1</th>
<th>16/1</th>
<th>17*/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>1,172,072</td>
<td>1,267,922</td>
<td>1,307,836</td>
<td>1,342,477</td>
<td>1,025,521</td>
<td>3.1</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>January</td>
<td>404,106</td>
<td>442,045</td>
<td>460,577</td>
<td>466,086</td>
<td>532,206</td>
<td>4.2</td>
<td>1.2</td>
<td>14.2</td>
</tr>
<tr>
<td>February</td>
<td>385,760</td>
<td>425,801</td>
<td>430,207</td>
<td>448,468</td>
<td>493,315</td>
<td>1</td>
<td>4.2</td>
<td>10</td>
</tr>
<tr>
<td>March</td>
<td>382,206</td>
<td>400,076</td>
<td>417,052</td>
<td>427,923</td>
<td></td>
<td>18.3</td>
<td></td>
<td>2.6</td>
</tr>
<tr>
<td>Q2</td>
<td>920,527</td>
<td>933,446</td>
<td>994,154</td>
<td>1,018,455</td>
<td></td>
<td>0</td>
<td>6.5</td>
<td>2.4</td>
</tr>
<tr>
<td>April</td>
<td>327,000</td>
<td>332,690</td>
<td>361,139</td>
<td>367,684</td>
<td>17.9</td>
<td>1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td>292,115</td>
<td>300,302</td>
<td>314,748</td>
<td>320,601</td>
<td>25.3</td>
<td>1.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>June</td>
<td>301,412</td>
<td>300,454</td>
<td>318,267</td>
<td>330,170</td>
<td>20.1</td>
<td>3.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>964,612</td>
<td>998,690</td>
<td>1,044,880</td>
<td>1,147,483</td>
<td></td>
<td>0</td>
<td>4.6</td>
<td>9.8</td>
</tr>
<tr>
<td>July</td>
<td>338,761</td>
<td>340,091</td>
<td>364,325</td>
<td>395,761</td>
<td>19.2</td>
<td>8.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>August</td>
<td>342,064</td>
<td>347,211</td>
<td>366,096</td>
<td>406,214</td>
<td>16.4</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>September</td>
<td>283,787</td>
<td>311,388</td>
<td>314,459</td>
<td>345,508</td>
<td>16.9</td>
<td>9.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>1,152,954</td>
<td>1,302,717</td>
<td>1,428,361</td>
<td>1,503,297</td>
<td></td>
<td>0</td>
<td>9.6</td>
<td>5.2</td>
</tr>
<tr>
<td>October</td>
<td>334,410</td>
<td>390,637</td>
<td>408,922</td>
<td>414,077</td>
<td>14.9</td>
<td>1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>November</td>
<td>386,737</td>
<td>411,501</td>
<td>444,640</td>
<td>477,686</td>
<td>16</td>
<td>7.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>December</td>
<td>431,807</td>
<td>500,579</td>
<td>574,799</td>
<td>611,534</td>
<td>12.9</td>
<td>6.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


A decade afterward of UNTAC intervention, international tourists reached up to around 700 thousand visitors in 2003, it rises almost 500% comparing to 1993. Currently, the trend is skewed slightly for more than 5 million visitors which equaled to one three of the Cambodian population. For instance, the industry is knowingly shocked by both
Asian crisis in 1997, the growth rate has diminished by 16% and in global crisis in 2008, has reduced by around 11%, (figure 2). Tourist receipts are augmented dramatically, accounted 3.2 million USD in 2016 as of that in 1995 attracted 1 million USD. Most of inbound tourists coming to Cambodia spent 6 to 7 days for staying with the pocket payment of 640$ per tourist (data calculated from the tourism statistics report, April 2017).

The strong commitment and policy base development of ministry of tourism and the related institutional state is putting in place for strong support. Well administrative and foreign policy toward China has brought Cambodia one of the most attractive place for China tourists. From 2010 to 2017, there are 7 million China tourists travelling to Cambodia. From which Cambodia is a home of both natural and cultural heritage, there are places to enjoy and leisure. Currently, the efforts have been made to attract additional arrivals by establishing more direct flights and introducing new initiatives such as the “China Ready” initiative and joint tour packages.

**Figure 1:** International tourist arrivals from 1993 to 2016

![Figure 1: International tourist arrivals from 1993 to 2016](image1)

**Source:** Tourism Statistics Report, Ministry of Tourism (2017)

**Figure 2:** Average length of stays and international tourism receipts, (1995 – 2016)

![Figure 2: Average length of stays and international tourism receipts, (1995 – 2016)](image2)

**Source:** Tourism Statistics Report, Ministry of Tourism (2017)
Estimation Outcomes from the Baseline Regression Equation

To estimate factors affecting tourist arrivals in Cambodia, the study employs monthly observations of international tourist arrivals as an explained variable, exchange rate as the main explanatory variable and some dummy variables such as the global financial crisis, the national election, AEC and Cambodia’s e-Visa as an exogenous variables. The empirical results show in table 4. With 209 of sample observations, all opposed models indicate the statistical significance with 1% level in line with F-statistic value. Thus, they are perfectly and correctly modified. From model (1) to (4), the study tests baseline regression through OLS with robustness SE whereas model (5) to (6) apply 2SLS with IV regression. Model (1) and (2) estimate the exchange rate only in the before-after the global financial crisis without controlling dummy variables. From model (3) to (6), the study uses exchange rate in the whole sample to estimate with and without dummy variables. In addition, model (5) imports all dummy variables as the instrumental variables and adopts 2SLS instrument variable regression.

It is reflected that exchange rate before the global financial crisis are negatively affected to TA with the statistical significance of 1%. This trend is showdown tourist arrivals resulted of rising exchange rate. Conversely, exchange rate after crisis and exchange rate in the whole sample, model (2) to model (6), are positively associated with TA. This is likely due to the facts that Cambodia has pegged its own currency to US dollar (USD), its appreciation and depreciation might not reflect strongly to tourist decision. Or meaning that Cambodia’s currency is allowed the value to be fluctuated due to supply and demand of market. The positive is given the facts of effecting and supporting the importance of maintaining a relatively stable exchange rate to attract international tourist arrivals. In addition, tourist arrivals to Cambodia is not sensitive to currency shock since Cambodia is able to maintain the appreciation and depreciation in the relative gap.
Table 4: Factor effecting tourist arrivals in Cambodia

<table>
<thead>
<tr>
<th>Tourist arrivals</th>
<th>Baseline Regression Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model (1)</td>
</tr>
<tr>
<td>Exchange rate</td>
<td>14.526***</td>
</tr>
<tr>
<td></td>
<td>(10.46)</td>
</tr>
<tr>
<td>Exchange rate x before-crisis</td>
<td>-1.27***</td>
</tr>
<tr>
<td></td>
<td>(-19.67)</td>
</tr>
<tr>
<td>Exchange rate x after-crisis</td>
<td>1.25***</td>
</tr>
<tr>
<td></td>
<td>(20.29)</td>
</tr>
<tr>
<td>Dummy variables</td>
<td></td>
</tr>
<tr>
<td>National election</td>
<td>0.138*</td>
</tr>
<tr>
<td></td>
<td>(1.97)</td>
</tr>
<tr>
<td>e-Visa</td>
<td>1.148***</td>
</tr>
<tr>
<td></td>
<td>(16.40)</td>
</tr>
<tr>
<td>Global financial crisis</td>
<td>-0.392***</td>
</tr>
<tr>
<td></td>
<td>(-5.96)</td>
</tr>
<tr>
<td>AEC</td>
<td>0.569***</td>
</tr>
<tr>
<td></td>
<td>(9.64)</td>
</tr>
<tr>
<td>Constant term (c)</td>
<td>12.55***</td>
</tr>
<tr>
<td></td>
<td>(363.23)</td>
</tr>
<tr>
<td></td>
<td>-108.83***</td>
</tr>
<tr>
<td></td>
<td>(-9.43)</td>
</tr>
<tr>
<td></td>
<td>-28.39***</td>
</tr>
<tr>
<td></td>
<td>-28.39***</td>
</tr>
<tr>
<td>Number of observations</td>
<td>209</td>
</tr>
<tr>
<td>F-statistics</td>
<td>386.77***</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
</tr>
<tr>
<td></td>
<td>109.51***</td>
</tr>
<tr>
<td></td>
<td>163.16***</td>
</tr>
<tr>
<td></td>
<td>84.25***</td>
</tr>
<tr>
<td></td>
<td>163.16***</td>
</tr>
<tr>
<td>Adjust R²</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>0.2843</td>
</tr>
<tr>
<td></td>
<td>n/a</td>
</tr>
</tbody>
</table>

Source: Author’s estimates

Note: Robust t statistics in parentheses and * p<0.05, ** p<0.01, *** p<0.001. OLS is estimated in line with robust standard error (SE). IV is estimated through instrument variable (IV) technique or so-called Instrumental variables (2SLS) regression. 2SLS (1) used all dummy variables to be an instrument variables. n/a defined not available information.

With respect to dummy variables, it is revealed the empirical outcomes as expected. The global financial crisis is negatively linked to tourist arrivals. It is likely occurrence to some empirical studies that a significant slowdown in the Turkish foreign active tourism during the global crisis (Kudret Gul 2014). In addition, (José F. P-R. et al. 2016) found that the proposed model is appropriate for explaining the changes in the market positions caused by the economic crises. According to (UNWTO 2013) stressed that the 2008–2009 global economic crisis severely affected international tourism, causing in 2009 a decline of 4% in international tourist arrivals and a decrease of 6% in international tourism receipts in 2009. The crisis actually caused the first serious downturn faced by international tourism in decades, a sector accustomed to a long-term average growth rate of about 4% a year. The World Bank also stated that the financial crisis has cut access to loans in advanced and developing countries, pulling investment out of poorer nations and reducing consumer spending. Hence, reducing consumption and investment, slowdown tourist travel, then.

Furthermore, Cambodia’s e-Visa and the AEC are found to be positively related to TA. Launching and adopting such the technology innovation brought Cambodia as an easy place to access to travel, particularly applying for short term Visa. This e-Visa seems to have been embraced. Certainly, it has reduced the time and expense required in securing official permission to travel to the sub-continent. Hence, Cambodia is one of the easiest countries in the world to emigrate to, visa-wise. The positive of the AEC is undoubtedly since AEC will bring not only trade and investment flow but service, labors as well as...
tourist arrivals with visa exemption for some countries within the nations. The study furthermore stressed that the national election is also positively associated to TA but it is shown a small proportion, meaning that it is not existed the strong impact. It is opposite to (Ghana | Bennet Otoo 2016) harassed that the word “elections” is often surrounded by a general stigma of fear, chaos, and anxiety. In every part of the world, the electioneering period is a brief period of a dip in almost every sector of life. A lot of activities are put on hold and investors/businessmen are reluctant to travel or do business in such countries at this time.

Estimation Outcomes from ARIMA and ARIMA-GARCH Models

To estimate and forecast tourist arrivals in line with TA variable individually, the study employs TA as the nature of the logarithm with the first different, I(1) to overcome the stationary process. Baseline regression between TA and TA at lag 1 is adopted to capture the pattern of TA at level, I(0) and first different data, I(1). The result shows in figure 4. It somehow defines that the simple regression with time trend effect and scatter plotting of sample data draw in the gap of 10 to 14 as converting to logarithm function and in the gap of -4 and 4 due to taking the first different data. TA at time (t-1) is positively and significantly associated to TA at time (t). It reflects that the more tourist arrivals in the past year, the more they visit in the present year. This gives the idea that tourists return to the country of visit due to level of pleasant and satisfying. According to research, there is 20% of tourist arrivals in Cambodia have been visited more than one time.

To apply Box-Jenkins methodology, detecting the random order and stationary process is though applied in the previous session, the study employs autocorrelation (AC) to determine the best fit parameter of AR and partial correlation (PAC) for MA parameter. As the result, it is shown in figure 5 as bellows.

**Figure 4:** Post-estimation in line with level and log differential data

Source: Author’s estimates

Figure 5 specifies that ACF contains significant values in the first three lags, while the PACF exhibits decay in the form of an approximate damped. As the result, parameter of AR could be tested by 1, 2, 3 and MA is tested from 1, 2, 3 and up to 4. It suggested as
an appropriate specification. The study selects the parameters of AR and MA due to graphing AC and PAC. The result suggests to adopt its parameter differently. Therefore, its empirical results show in table 5. The study proposes a post estimation from univariate time series models, namely ARIMA, GARCH and hybrid of ARIMA-GARCH model with the interaction of time trend (t). It employs full sample of 2000m1 to 2017m7, covering 208 observations.

**Figure 5:** Autocorrelation (AC) and partial correlation (PAC) at log different data

![Figure 5: Autocorrelation (AC) and partial correlation (PAC) at log different data](image)

**Source:** Author’s estimates

Table 5 shows that constant term (c) and time trend is still insignificant and keep the sign constantly. More interestingly, it is such a straight forward that constant term (c) is insignificant. It is not a must to drop the constant term in the models3. Most of AR and MA coefficients are negatively and positively affected to TA. In model (1), say ARIMA (1, 1, 1) demonstrates that MA(1) is positively affected to TA whereas AR(1) reveals insignificant association. The positive of MA reflects the statistical method that tourist arrivals (TA) contains an MA term in the model. Coefficient of AR and MA of model (2), (3) and (4) show the significant relationship to TA. Coefficient of AR(1) is positively affected where AR(2) and AR(3) are negatively related to TA. The positive relationship of AR(1) to TA reflects the facts of tourist arrival at the present time, say time (t) is impacted by tourist at past time, say (t-1) but it is reduced by time destructive. It means that when time trend is reduced due to AR coefficient, tourist arrivals will reduce as it is shown a negative impact, AR(2) and AR(3).

Therefore, the most suitable model to adopt an out of sample forecasting is model 4, say ARIMA (3, 1, 4) due to the lowest statistical value of AIC and BIC and significant level of the Wald chi-square, which is approximated 6.49. Yet, ARIMA (3, 1, 4) could be considered to model with GARCH (1, 1) as the hybrid model, say ARIMA (3, 1, 4)-GARCH (1, 1). Consequently, ARIMA (1, 1, 1)-GARCH (1, 1) and model 7 takes into account. From model (5) and (6), the study applies GARCH (1, 1), ARIMA (3, 1, 4)-GARCH (1, 1) and ARIMA (1, 1, 1)-GARCH (1, 1). It shows that AR(3) remains negatively associated with TA where AR(1) is not significant impact. In line with ARCH and GARCH coefficient specify only β is significant impact. The insignificant of α reveals that volatility shock today of tourist arrivals is not fed through into next period’s volatility. More importantly, GARCH(1, 1) model of tourist arrivals to

---

3 Usage Note 23136: Understanding an insignificant intercept and whether to remove it from the model, URL: [http://support.sas.com/kb/23/136.html](http://support.sas.com/kb/23/136.html)
Cambodia suggests that the short run persistence of shocks lies in the gap of 0.04 while the long run persistence lies in the gap of 0.94. As the second moment condition, $\alpha + \beta = 0$ is satisfied.

Table 5: Tourist volatility, estimation of full sample observations

<table>
<thead>
<tr>
<th>Models</th>
<th>ARIMA (p, d, q)</th>
<th>GARCH (s, r)</th>
<th>ARIMA (p, d, q)-GARCH (s, r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) ARIMA (1, 1, 1)</td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>(2) ARIMA (1, 1, 4)</td>
<td>(0.07)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>(3) ARIMA (2, 1, 4)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>(4) ARIMA (3, 1, 4)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>(5) GARCH (1, 1)</td>
<td>(0.52)</td>
<td>(0.58)</td>
<td>(0.42)</td>
</tr>
<tr>
<td>(6) ARIMA (1, 1, 4)-GARCH (1, 1)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>(7) ARIMA (1, 1, 4)-GARCH (1, 1)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
</tbody>
</table>

Constant term: 
(1) $0.06$ (0.55) 
(2) $0.07$ (0.46) 
(3) $0.05$ (0.76) 
(4) $0.06$ (0.59) 
(5) $0.05$ (0.52) 
(6) $0.06$ (0.58) 
(7) $0.05$ (0.42)

Sigma: 
(1) $0.15***$ (17.62) 
(2) $0.15***$ (16.97) 
(3) $0.15***$ (18.48) 
(4) $0.15***$ (17.12) 
(5) $0.15***$ (17.62) 
(6) $0.15***$ (16.97) 
(7) $0.15***$ (18.48)

AR parameters: 
(1) $\phi_1$ = -0.2 (-0.77) 
(2) $\phi_2$ = -0.28*** (-3.70) 
(3) $\phi_3$ = -0.13** (-1.79) 
(4) $\phi_4$ = 0.45** (1.93) 
(5) $\phi_4$ = 0.25*** (-1.00) 
(6) $\phi_4$ = 0.13*** (1.94) 
(7) $\phi_4$ = 0.13** (1.94)

MA parameters: 
(1) $\psi_1$ = 0.04 (0.93) 
(2) $\psi_1$ = 0.0004 (0.00) 
(3) $\psi_1$ = 0.04 (0.92) 
(4) $\psi_1$ = 0.0004 (0.00) 
(5) $\psi_1$ = 0.0004 (0.00) 
(6) $\psi_1$ = 0.0004 (0.00) 
(7) $\psi_1$ = 0.0004 (0.00)

ARCH & GARCH parameters: 
(1) $\omega$ = 0.94*** (12.16) 
(2) $\omega$ = 0.19 (12.76) 
(3) $\omega$ = 0.94*** (12.16) 
(4) $\omega$ = 0.19 (12.76) 
(5) $\omega$ = 0.0002 (0.19) 
(6) $\omega$ = 0.0177 (0.00) 
(7) $\omega$ = 0.0004 (0.00)

Testing Parameter Coefficient in GARCH (r, s): 
(1) $\alpha + \beta = 0$ 
(2) $\alpha + \beta = 0$ 
(3) $\alpha + \beta = 0$ 
(4) $\alpha + \beta = 0$ 
(5) $\alpha + \beta = 0$ 
(6) $\alpha + \beta = 0$ 
(7) $\alpha + \beta = 0$

AIC: 
(1) -199.9 
(2) -200.92 
(3) -199.26 
(4) -193.47 
(5) -193.36 
(6) -189.52 
(7) -193.36

BIC: 
(1) -183.22 
(2) -184.24 
(3) -182.57 
(4) -176.78 
(5) -176.67 
(6) -166.16 
(7) -178.22

Wald chi2: 
(1) 21.58*** 
(2) 19.11*** 
(3) 14.95*** 
(4) 6.49* 
(5) 0.17 
(6) 6.07** 
(7) 18.10***

Observations: 
(1) 208 
(2) 208 
(3) 208 
(4) 208 
(5) 208 
(6) 208 
(7) 208

Source: Author’s estimates

Note: Model 1, 2, 3 and 4 refer to ARIMA (1, 1, 1), ARIMA (1, 1, 4), ARIMA (2, 1, 4) and ARIMA (3, 1, 4) respectively. Model 5, 6 and 7 denote GARCH (1, 1), ARIMA (3, 1, 4) – GARCH (1, 1) and ARIMA (1, 1, 1) – GARCH (1, 1) respectively. The sign notification of *, ** and *** referred to the statistical significance of 10%, 5% and 1%.

Estimation Outcomes of Measurement Predictive Accuracy

In this session, we aim at measuring the forecasting error from out of sample prediction due to the lowest value of RMSE and U index. This adopts the appropriated models throughout ARIMA (p, d, q) and GARCH (s, r) with one step (1-step) ahead obtained from table 5. The first 1-step ahead applies the post estimation from sample observations of 2000m1 to 2013m12. Forecasting out of sample is afterward considered from 2013m12 to 2017m7. The second 1-step ahead starts from 2014m12, the third 1-step ahead from 2015m12 and the fourth 1-step ahead from 2016m12 to 2017m7, respectively. As the result, the statistical value of RMSE and U report in table 6 and the comparison of the models illustrates in figure 6. It demonstrates that from the model (1)
to model (5), RMSE and U index is beyond and insight into 1-step ahead as it produced the smallest error amongst others. It is likely due to the facts that the models are perfectly fitted with long period ahead rather than the nearest period.

**Figure 6: One-step prediction of residuals from the post estimation**

![Figure 6: One-step prediction of residuals from the post estimation](image)

Source: Author’s estimates

More importantly, to capture the gap of forecasting error from the best modified model, the study employs the quantile regression model in line with different conditional distributions, say 25%, 50%, 75% and 95%. This guide to analyze the fitted and the actual value of the sample toward the residual of forecasting. Since the long term period forecasting from 2013m12 to 2017m7 produced the smallest error due to the statistical value of RMSE and U index, the study applies this period to predict residuals at different conditional quantile. As the result, figure 7 shows that point forecasting interval of quantile at 25% and 50% fits perfectly to the actual value and shows the errorless rather than the others.

**Table 6: Measurement predictive results from out of sample forecasting with n-step ahead**

<table>
<thead>
<tr>
<th>Models</th>
<th>2013m12 – 2017m7</th>
<th>2014m12 – 2017m7</th>
<th>2015m12 – 2017m7</th>
<th>2016m12 – 2017m7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>Theil's U</td>
<td>RMSE</td>
<td>Theil's U</td>
</tr>
<tr>
<td>(1)</td>
<td>0.1386</td>
<td>0.8801</td>
<td>0.1471</td>
<td>0.9562</td>
</tr>
<tr>
<td>(2)</td>
<td>0.1322</td>
<td>0.7777</td>
<td>0.1401</td>
<td>0.9498</td>
</tr>
<tr>
<td>(3)</td>
<td>0.1338</td>
<td>0.8972</td>
<td>0.1422</td>
<td>1.0534</td>
</tr>
<tr>
<td>(4)</td>
<td>0.1394</td>
<td>0.8884</td>
<td>0.1484</td>
<td>0.9416</td>
</tr>
<tr>
<td>(5)</td>
<td>0.1325</td>
<td>0.8119</td>
<td>0.1402</td>
<td>0.9458</td>
</tr>
</tbody>
</table>

Source: Author’s estimates

Note: model 1 is ARIMA (1, 1, 1), model 2 is ARIMA (3, 1, 4), model 3 is GARCH (1, 1), model 4 is ARIMA (1, 1, 1) – GARCH (1, 1) and model 5 is an ARIMA (3, 1, 4) – GARCH (1, 1). All models were displayed by time trend effect (t).
**CONCLUDING REMARKS**

International tourist arrivals are the crucial source of revenue for many developed and developing countries. Therefore, to predict its trend is ideal and to manage international tourism growth is essential and compulsory to model and forecast adequately tourist arrivals and their associated volatility with its order of parameters. This study aims at modeling and forecasting tourist arrivals in Cambodia from monthly observations of 2000m1 to 2017m7, covering 209 samples. The empirical results show that in one part toward the baseline regression equation, tourist arrivals have a significant occurrence resulted of the appreciation of exchange rate and some internal and external factors such as the national election, the AEC, e-Visa application as well as the financial global crisis. These factors appearance the power explanation with a statistical significance during the observed periods. In another part regarding to autoregressive and volatility approach, the empirical results indicate that tourist arrivals is affected by time trend and the previous visitors in the past period, say time (t-1). More importantly, the trend is reduced due to time lag, say time (t-2, t-3, t-4). The GARCH (1, 1) model of tourist arrivals suggests that the short run persistence of shocks lies in the gap of 0.04 while the long run persistence lies in the gap of 0.94. RMSE and U index obtained from the measurement predictive accuracy of out of sample forecasting reveal that long run 1-step ahead of the period 2013m12 to 2017m7 is produced the smallest error among the others. Thus, it has more predictive power to apply long term ex-ante forecasting.

Nonetheless, the empirical findings print out some messages and suggestions for the further academic researcher as well as policy makers. Tourism policy makers should consider carefully the unexpected events which cause the volatility of tourist arrivals, say the economic crisis e.g., and national event such as national election in line with some sensitive factors that might disturb strongly to travel decision. In addition, once the volatility of tourist arrivals is found, it does matter for tourism and business policy makers whether which path of the shock the generating policies to attract tourists could be employed effectively. So far, the tourism sector is a relevant economic activity, it is
significant and necessary to note the unanticipated shock in line with volatility shock, will have an inference on tourist arrivals for Cambodia, both in the short and long run. Alternatively, it is necessary to determine the extent to which a volatility shock, the tourism is diverted to other countries that particularly have similar product development of tourism industry. The facts that economic and political events may be changed and occurred in the real and exact economic phase with some interaction of an exogenous factor.

REFERENCES


