Online Shopping Behavior in Thailand

Inpong Luanglath\textsuperscript{1} and Panita Khongtone\textsuperscript{2}
Bangkok University – International College, Thailand
\textsuperscript{1}lecturepedia@gmail.com
\textsuperscript{2}panita.khon@bumail.net

Abstract - The objective of this paper is to test whether online shopping behavior depends on multiple factors, namely access to online shopping platform, intent to shop online, attitude and perception of the e-commerce infrastructure. Through multiple regressions, the weight of each factor may be determined. With known weight of each influencing factors, stakeholders could better plan and execute e-commerce business. The purpose of this research is to provide a clearer picture of online shopping behavior among online consumers in Thailand. The data used for this research comes from field survey of 50 samples. Quantitative method involving continuous and discrete probabilities were used as analytical tools. Parametric and non-parametric tests were used to determine data characteristics and modeling. It was found that under simple regression, access to the online platform (X1) is not statistically significant. Other factors, such as intent to shop online (X2), attitude on online shopping (X3) and the importance e-commerce infrastructure (X4) are significant with T score of 1.76, 1.74 and 2.28 respectively compared to the theoretical value of 1.64. Under multiple regression, X2 and X4 are significant contribution to online shopping behavior where F = 2.84 compared to the theoretical value of 1.69. The interaction effect test shows that access to online platform is a significant factor: $\beta_3X_2X_4$ of T = 1.84. The intended contribution of this research is to provide stakeholders with practical model for online shopping behavior from the perspective of online consumers. With known behavior, stakeholders could better put e-commerce in Thailand in perspective.

Keywords - Confirmatory Factor Analysis (CFA), Goodness of Fit, Interaction Effect, Likelihood Estimate, Online Shopping

I. INTRODUCTION

Thailand has a population of 67 million people; a survey in 2015 showed that 23.9 millions or 37% of the total population are active internet users. Among these numbers, 17.7 millions or 27% of the total population are active mobile internet users. These internet users are target population for e-commerce. The same survey showed that 19% of internet users use PC for online search and 18% resulted in online purchase. Similarly, 12% used mobile phone for online search and 11% resulted in online purchase. Although e-commerce or online shopping in Thailand currently does not command a significant size, but the potential growth for e-commerce in Thailand is not negligible. There are approximately 97 million mobile phone subscribers. With such a market potential, online shopping behavior of people in Thailand is of interests to stakeholders, such as researcher, web store owners, policy makers, and online shoppers.

The conceptual framework of this research is to construct a model that would explain online shopping behavior from the perspective of consumers. The proposed model defines the dependent variable (Y) as online shopping behavior. The independent variables include access to online shopping platform (X1), intent to purchase (X2), attitudes towards online shopping (X3) and perception of e-commerce infrastructure in Thailand.
The seller’s perspective on online shopping is not within the scope of this research. Most literature focuses on web store or seller’s perspective and to make an inference about online shoppers’ behavior. This research attempts to define online shoppers’ behavior from empirical data by surveying internet users. The survey asks about their attitude and perception towards online shopping in Thailand.

II. LITERATURE REVIEW

A. Modeling Approach in Social Science Research

In general, there are three types of modeling approach; these include exploratory factor analysis (EFA), confirmatory factor analysis (CFA) and structural equation modeling (SEM). EFA involves the exploration of possible explanatory factors not yet covered by existing literature [42], [18]; and [22]. CFA involves the use of explanatory factors covered in the current literature. The purpose of the test in CFA is to confirm or disaffirm current theory [46], [32] and [11]. Lastly, in SEM the variables involve general subject matter and the data distribution behaves in predictable matter so as to allow generalizability [27], [37] and [50]. This research employs CFA modeling by using Thailand as a case study to test whether explanatory factors claimed by the current literature would still be valid for online shopping behavior in Thailand. The conceptual framework of CFA is illustrated by Fig. 1 below.

Online shopping is an e-commerce activity. E-commerce is defined as the exchanges of goods and services via the Internet. E-commerce is an important tool for business worldwide [15]. E-commerce enhances the company’s competitiveness [4] by allowing access to world wide market, increase customer value and sustainable capabilities. From seller’s perspective, a successful e-commerce model depends on three factors: quality of information system, services, and user satisfaction [14]. In order to maintain their competitiveness, firms must continually define, design, and develop web store to meet the demand of increasingly sophisticated online shoppers [19].

Web designers must research consumer expectation. Consumer expectation in online shopping goes beyond user-friendly website. Web stores must be able to assure online shoppers data security. Factors that lead to consumer shopping online are convenience, selection, competitive pricing, and access to information [31] and [43]. This research assesses e-commerce in Thailand through the perception and behavior of online shoppers. The stakeholders for whom the utility of this research are aimed include web store owners, online shoppers, policy makers and researchers looking to explore the e-commerce market.
potential in Thailand.

**B. Consumer Decision Making as the basis for Online Shopping Behavior**

Consumer decision making models may be explained in several perspectives [40], [16] and [30]. Current literature in consumer behavioral research categorized these perspectives into four lines of approaches. The first line of literature on consumer decision making focuses on personality perspective of decision making [25]. Individuals have different personalities. Some are introverts and others are extraverts [17]. These differences lead to differences in decision making process [38].

The second line of literature looks at motivational perspective of decision making. Psychological attributes, such as motivation make lead to differences in decision making among consumers. These differences in motivational influence for decision making led to the use of segmentation of consumers [26]. The source of such motivation may come from lifestyle, interests or opinions [53]. Motivation is the basis for intent to engage in a transaction. This second line of the literature provides support for our use of the second IV (X2 = intent for online shipping).

The third line of literature studies the attitude perspective of decision making among consumers. Attitude is the feeling or belief of the consumer prior to entering into the transaction. Past experience could influence attitude [45]. Attitude may result from experience with multiple stimuli [23]. This third line of literature provides support for our third IV (X3 = attitude) as a possible explanatory factor.

Lastly, the fourth line of literature claims that the situation surrounding the transaction influences the consumer’s decision making. Situation may shape a consumer’s behavior, experience, and perception [34]. For instance, the theory of Gestalt psychology posits that the “totality of experience” shapes the consumer’s decision making [35]. The collection of these experiences allows the consumer to make sense (subjectively) of the multiple stimuli [10]. Consequently, the consumer prioritizes what to buy and not to buy [21]. The situation created by online shopping is the online experience. From this experience, online shopper may formulate a perception towards online shipping. This line of research provides support for our measurement of the consumer perception towards e-commerce infrastructure in Thailand (X4).

**III. DATA**

Primary data were used for this research. Written survey was used as the instrument to collect the data. The dependent variable (Y) was defined as online purchasing behavior. Independent variables (Xs) include: access to online platform (X1), intent (X2), attitude (X3), and perception of internet infrastructure in Thailand (X4). The survey solicits three types of data, namely quantitative, ordinal and nominal data. Quantitative was obtained through the used of (0, 1, 2, 3) scale. The scale was calibrated for item response reliability by:

\[
R = \frac{e^{\alpha}}{100} + \sqrt{1 - df(\alpha)}
\]

where \(e = 2.718\), \(\alpha = \) error level, and \(df = \) degree of freedom or \(df = n - 1\). In the present case, the number of choice in the response is \(n = 4\) where 0 = none, 1 = low, 2 = medium, and 3 = high. The degree of freedom is \(df = 3\). Using 0.95 confidence interval, the error level is \(\alpha = 0.05\). The calibrated prospective reliability for the scale is 0.9491 or about 94.91% compared to the Likert scale of (1, 2, 3, 4, 5) which has item response reliability of 0.8931.

Minimum sample size determination was accomplished by the \(n_\Omega\) (n-Omega) method [36]. The \(n_\Omega\) method is given by:

\[
 n_\Omega = \sqrt{\left(\frac{n_3}{0.01}\right) + \left(\frac{n_3}{0.99}\right)}
\]

where \(n_3\) is the chosen number of stimuli.
where \( n_3 = \sqrt{n_1 - n_2}/2 \), \( n_1 = (Z\sigma)/E \), \( n_1 = (Z^2\sigma^2)/E^2 \), \( Z = \) critical value for the unit normal distribution \( Z \) at a specified confidence interval, \( \sigma = (\bar{x} - \mu)/Z \sqrt{n} \), and \( E = \sigma/\sqrt{n} \). The terms \( \bar{x} \) and \( \mu \) are the test-sample mean and the estimated mean for the test sample. In this study, test-samples of 10 counts were used. Repeated measurements of three test-samples were employed. The minimum sample required for field survey was \( n_\Omega = 30 \). We collected 50 surveys from the field.

The sample frame consists of general population in the Bangkok metropolitan. Since the population size for the target population is non-finite, it is not possible to implement random sampling because random sampling requires that each element in the population has equal chance of being selected [48]. In order to determine the selection probability for each element, a finite population is required. This requisite for randomness was absent in the target population. However, for linear regression, it is sufficient to show that the residuals are statistically independent and uncorrelated, i.e. \( \text{Cov}(\varepsilon_i, \varepsilon_j') = 0 \forall i \neq j' \) [9].

### IV. METHODOLOGY

The sample was tested for data characteristics before it was subjected to modeling and fitting. The data characteristic tests includes distribution test (Table II). A distribution test was used to verify that all variables are similarly distributed before they are fitted into regression modeling.

Data distribution test is accomplished by Anderson-Darling (AD) [2] and [3]. The observed value for AD is obtained through:

\[
AD = -n - S
\]  

where \( S \) is defined as:

\[
S = \sum_{i=1}^{n} \left( \frac{2k-1}{n} \right) \left[ \ln(F(Z)) + \ln(1 - F(Z)) \right]
\]

The theoretical value for AD* is given by:

\[
AD^* = AD \left( 1 + \frac{0.752}{n} + \frac{2.25}{n^2} \right)
\]  

The decision rule is: \( H_0 : AD < AD^* \) assumed that the data is normally distributed and \( H_A : AD > AD^* \) that the data is non-normally distributed.

#### A. Data Randomness Test

Randomness test (Table III) was used to verify that all discrete and continuous data are random events. Data randomness is a basic requirement for general statistical test using T and Z tables [49]. Furthermore, in regression analysis, dependent variable (DV) and independent variable (IV) must be random, i.e. independent and identically distributed (i.i.d.) [13]. Randomness test for quantitative data was accomplished by the adjacent test. For \( n < 25 \), the adjacent test is given by:

\[
L_{n<25} = \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

Out of 50 surveys collected, 25 items were selected for the adjacent test. The null hypothesis assumes randomness and this assumption may be rejected if the test statistic lies outside the lower and upper bounds of the critical value. The hypothesis statement is: \( H_0 : L_{\text{lower}} < L_{\text{obs}} < L_{\text{upper}} \) and the data is non-random if \( L_{\text{obs}} \) lies outside of the lower and upper boundaries.

Discrete date may be tested for randomness under the NIST’s monobit method:

\[
S_{\text{obs}} = \frac{|S_n|}{\sqrt{n}} \text{ where } S_n = \sum_{i} (+,-)_{i}
\]

The observed value \( S_{\text{obs}} \) is compared to the theoretical value:
Inpong Luanglath and Panita Khongtone

\[ Z_{\text{mono}} = \frac{S_{\text{obs}}}{\sqrt{2}} \]  

(8)

The decision rule is governed by:

- \( H_0 : Z \leq 1.65 \) for randomness, otherwise \( H_A : Z > 1.65 \) which means that the data set is not random [47].

**B. Extreme Value and Significant Outlier Test**

To assure that the survey does not contain any defects resulted from extreme values, generalized extreme value (GEV) analysis was used to determine whether significant outliers exist (Table IV). Extreme values would inflate error and undermine the accuracy of the model. If extreme values exist, they must be removed.

Extreme values may be analyzed under the generalized extreme value (GEV) distribution proposed by Fisher-Tippett-Gnedenko [24]:

\[ H(x; \mu, \sigma, \xi) = \exp \left\{ -\left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \]  

(9)

where \( \mu = \) location; \( \sigma = \) scale; and \( \xi = \) shape. If \( \xi > 0 \), \( H \) becomes a cumulative distribution function (CDF); if \( \xi < 0 \), it is valid for \( x < \mu + \sigma \xi \); and if \( \xi = 0 \), \( H \) is undefined [6]. However, if \( \xi \to 0 \), then (9) is reduced to:

\[ H(x; \mu, \sigma, 0) = \exp \left\{ -\left( \frac{x - \mu}{\sigma} \right) \right\} \]  

(10)

The parameter \( \xi \) is the tail index of the distribution. This index is used to classify the type of extreme value distribution. If \( \xi = 0 \), the \( H \) distribution is Gumbel distribution, also known as Type I where \( x \in \mathbb{R} \) and \( \xi = 0 \). The Gumbel distribution is given by:

\[ H(x; \mu, \sigma, 0) = \exp \left\{ -\exp \left( \frac{x - \mu}{\sigma} \right) \right\} \]  

(11)

If \( \xi > 0 \), the \( H \) distribution is a Fréchet distribution or Type II. The Fréchet distribution is given by:

\[
H(x; \mu, \sigma, \xi) = \begin{cases} 
0 & \text{for } x < \mu \\
\exp \left( \left( \frac{x - \mu}{\sigma} \right)^{-\alpha} \right) & \text{for } x > \mu 
\end{cases}
\]  

(12)

There are two methods for the tail index estimation: the Pickands method [44], and the Hill method [51]. The Pickands method is given by:

\[
\hat{\xi}_{k,m} = \frac{1}{m} \sum_{i=1}^{k} \left( \ln X_{n-i+1} - \ln X_{n-m} \right)
\]  

(13)

where \( m = \) number of observations whose tail is to be observed and \( k = \) sample size. The Hill method is given by:

\[
\hat{\xi}_{k,T} = \frac{1}{k} \sum_{i=1}^{k} \left( \ln R_{i,T} - \ln R_{k,T} \right)
\]  

(14)

where \( R = \sigma Z \); recall that \( \sigma \) is the estimated population standard deviation and \( Z \) is the standard score of the series. Both methods follow the same rule for classifying the type of extreme value distribution: Fréchet \( = \xi > 0 \), Weibull \( = \xi < 0 \) and Gumbel \( = \xi = 0 \).

**C. Central Limit Theorem Testing as Requisite for Regression Modeling**

Before the data was fitted in regression model, they are tested for the Central Limit Theorem (Table V). CLT serves several functions. CLT verifies whether the sample size was adequate. If the sample size is adequately large, the data will be normally distributed. CLT is manifested in normally distributed data. Secondly, CLT also verifies i.i.d. among DV and IVs in the proposed model. The Lindeberg-Lévy CLT method was used for this verification [5] and [8]:
Online Shopping Behavior in Thailand

\[ \lim_{n \to \infty} \Pr \left[ \sqrt{n}(\bar{X} - \mu) \leq Z \right] = \Phi \left( \frac{z}{\sigma} \right) \] (15)

where \( n \) = sample size; \( \bar{X} \) = sample mean, \( \mu \) = estimated mean, \( z \) = critical value at a specified percentage confidence interval read from the \( Z \)-Table, and \( \sigma \) = estimated standard deviation obtained via \( \sigma = \left[ \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{n}} \right] \). For non-standard normal distribution, the Lindeberg-Levy CLT may be written as:

\[ \left[ \sqrt{n}(\bar{X} - \mu) \right] / \sigma. \]

D. Simple and Multiple Regression Modeling

Each IV is regressed against the DV in simple regression to determine whether each IV significantly explains the occurrence of the DV (Table VI). For all IV variables that are statistically associated with DV, they are put together and modeled in multiple regression (Table VII). Interaction effects among IV's were also measured to determine whether the reading of DV is inflated due to the interaction among one or more of the IV pairs (Table VIII). Recall that in this study \( Y = \) online shopping behavior, \( X_1 = \) access to online platform, \( X_2 = \) intent to shop online, \( X_3 = \) attitude about online shopping, and \( X_4 = \) perception on e-commerce infrastructure in Thailand.

A simple regression model is given by:

\[ Y = \beta_0 + \beta_1 X_1 + error \] (16)

where \( \beta_0 \) = intercept, \( \beta_1 \) = parameter or coefficient to be estimated, and \( X_1 \) = independent variable. The test of statistical significance for simple regression is given by:

\[ T_r = \frac{r \sqrt{n - 2}}{\sqrt{1 - r^2}} \] (17)

where \( r = b(S_x / S_y) \) or correlation coefficient and \( n \) = sample size.

Multiple linear regression involves the use of two or more IV; for 4 explanatory factors, the model is given by:

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + error \] (18)

The test statistic is given by the analysis of variance (ANOVA) under the F-test:

\[ F = \frac{MSR}{MSE} \] (19)

where …

\[ MSR = SSR / p \] (19.1)
\[ SSR = \sum (\hat{Y} - \bar{Y})^2 \] (19.2)
\[ MSE = SSE / (n - p - 1) \] (19.3)
\[ SSE = \sum (Y_i - \hat{Y})^2 \] (19.4)

The interaction effect among IVs may be determined by:

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \ldots + \beta_p X_p + error \] (20)

where …

\[ \beta_3 = \hat{\beta} - \beta_1 - \beta_2 - \beta_0 \] (20.1)
\[ \hat{\beta} = \bar{Y} - (\beta_2 - \beta_0) - \beta_1 \] (20.2)

The term \( \beta_3 X_1 X_2 \) represents the interaction between \( X_1 \) and \( X_2 \). In similar fashion, the interactions among \( X_1 X_3, X_1 X_4, X_2 X_3, X_2 X_4 \) and \( X_3 X_4 \) may be calculated. The test statistics for the interaction effect is given by:

\[ T_i = \frac{\tilde{\beta}_1 - \tilde{\beta}_2}{\sqrt{\frac{1}{n_1} \frac{SE_1^2 + n_2 SE_2^2}{n_1 + n_2 - 2}}} \] (21)

If the value of \( T \) is less than 1.64, the interaction term is removed. The model is reduced to the original form as shown in (18).


E. Correlation Coefficient Testing according to Data Types

For discrete data testing, various correlation coefficients were tested according to their data types crossing between IV and DV (Table IX). These tests allow us to verify whether segmentation is appropriate for online shopping behavior. The use of quantitative, ordinal and nominal data mixing among the IV and DV provides a clearer picture of the relationship among the variables.

F. Goodness of Fit Tests to Verify Explanatory Power of the Proposed Model

The proposed model was tested for goodness of fit. The goodness of fit allows us to evaluate whether the proposed model is adequately robust to explain the data. The goodness of fit tests were achieved through the use of (i) likelihood ratio, (ii) Wald statistic, and (iii) Lagrange multiplier. The goodness of fit test was achieved through the use of eta ($\eta$) and the observed value. Eta is obtained via the QQ plot steps as follows: (i) determine $X_q$ and $Y_q$, (ii) run regression of $X_q$ and $Y_q$ to obtain the intercept for the line, and (iii) obtain the value for $\eta = \exp(a)$ where $a$ is the intercept (Appendix 2).

G. Log-Likelihood Ratio Test for Goodness of Fit

The likelihood ratio test is based on chi square distribution with degree of freedom: $df = df_2 - df_1$. The ratio calculation is the likelihood of the null divided by the likelihood of the proposed model. The test statistic is given as $\Lambda(x)$ as:

$$\Lambda(x) = \frac{L(\theta_0 \mid X)}{L(\theta_1 \mid X)}$$

or equivalently:

$$\Lambda(x) = \frac{L(\theta_0 \mid X)}{\sup\{L(\theta \mid X) : \theta \in \{\theta_0,m\theta_1\}\}}$$

where $L(\theta \mid X)$ is likelihood function, “sup” is the supremum function. The decision rule is governed by if $\Lambda > c$ do not reject the null hypothesis and if $\Lambda < c$ then reject the null hypothesis. The rejection point is the probability $\Lambda = c$. The variable $c$ and $q$ are selected at specified alpha (error) level whose relationship may be summarized as: $qP(\Lambda = c \mid H_0) + P(\Lambda < c \mid H_0) = \alpha$. The likelihood ratio test is a tool against Type I error. Type I error occurs when the null hypothesis is wrongly rejected. The likelihood ratio test has been classified as a power test [39]. Equations (24) and (25) are rewritten as:

$$\Lambda(x) = \frac{\sup\{L(\theta \mid x) : \theta \in \theta_0\}}{\sup\{L(\theta \mid x) : \theta \in \theta\}}$$

The calculation for the likelihood ratio follows a chi square hypothesis testing. With a sample of 50 surveys, the ratio is 1 or near one for all markets. This near 1.00 result shows that the estimation is close to the actual observed value or arithmetic mean. The critical value against which the ratio is test is 55.80.

H. Wald Statistic Test for Goodness of Fit

The third test to assess the likelihood function is the Wald statistic. For a single-parameter scenario, the Wald statistic is given by:

$$W = \frac{(\hat{\theta} - \theta_0)^2}{\text{var}[\hat{\theta}]}$$

This test is compared to the chi square in case where the data distribution is not normal. In case where the data is normally distributed, the Wald test is given by:

$$W_N = \frac{\hat{\theta} - \theta_0}{se(\hat{\theta})}$$

where $se$ is the standard error of the MLE estimate which is given by:

$$se = \frac{1}{\sqrt{I_n(MLE)}}$$
where $I_n$ is the Fisher information [28], [16] and [1]. The finding of the Wald test in Table 10 shows that there is no significant difference between the arithmetic mean and the estimated mean. The practical implication for online shopping behavioral analysis is that the new LLE can provide a more accurate estimation.

I. Langrange Multiplier Test for Goodness of Fit

The Langrange multiplier test is also called the score test. The score test had been explained by several authors, such as [7], [16], and [12]. The score test is more appropriate where the deviation between $\hat{\theta}$ and $\theta$ is small. The score test is given by:

$$U(\theta) = \frac{\partial \log L(\theta | X)}{\partial \theta}$$

The null hypothesis is $\theta = \theta_0$. If the null hypothesis cannot be rejected, the data is treated as chi square distribution. The test statistic is given by:

$$S(\theta_0) = \frac{U(\theta_0)^2}{I(\theta_0)}$$

where $I(\theta_0)$ is the Fisher information or

$$I(\theta_0) = -E \left[ \frac{\partial^2}{\partial \theta^2} \log L(X | \theta) | \theta \right]$$

For normally distributed data, the score test is given by:

$$S^*(\theta) = \sqrt{S(\theta)}$$

V. FINDINGS

As part of the preliminary tests of data characteristics, the Anderson-Darling test was used to verify data distribution for all variables. It was found that for both IV and DV, the data were normally distributed. For online shopping behavior study, this finding or normality among the variables implies that the relationship between $Y$ and $X_1$, $X_2$, $X_3$, and $X_4$ may be modeled under linear regression. Table II delineates the empirical and theoretical values for the AD test.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Quantitative Data</th>
<th>Null Hypothesis $AD &lt; AD^*$ Normal</th>
<th>Alternative Hypothesis $AD &gt; AD^*$ Non-normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td>50.62</td>
<td>51.89</td>
<td></td>
</tr>
<tr>
<td>$X_1$</td>
<td>58.87</td>
<td>60.34</td>
<td></td>
</tr>
<tr>
<td>$X_2$</td>
<td>51.33</td>
<td>52.62</td>
<td></td>
</tr>
<tr>
<td>$X_3$</td>
<td>46.10</td>
<td>47.25</td>
<td></td>
</tr>
<tr>
<td>$X_4$</td>
<td>49.74</td>
<td>50.99</td>
<td></td>
</tr>
</tbody>
</table>

The second test was to verify whether the data are random. Most statistical tests require that the data be random or came from a random process. The fact that the data was randomly selected does not guarantee that the data would be randomly distributed. Additionally, the fact that the data was not collected from a non-random sampling method does not vitiate the randomness of the data distribution.

Two types of data were used for randomness testing. Continuous data from both IV and DV were tested for randomness under the adjacent test. The range of value for the null hypothesis to find randomness is $1.37 < L < 2.63$. Under this decision rule, all values for the variables fall within the range of the null hypothesis. All quantitative data for IV and DV are found to be random.

These findings imply that the data came from a random process. Even though the survey was collected from purposive sampling method, the fact that the adjacent and run tests confirmed that the data came from a random process attests that the survey response represents un-manipulated responses.

For online shopping behavior study, the finding of randomness for the IV and DV series implies that online shoppers acted independently in providing their opinions and perception on online shopping. Moreover, a
finding of randomness in the data implies that there is no predictable pattern in any particular issues asked in the survey. Table III delineates the empirical results for the randomness test for continuous and discrete data in IV and DV series.

### TABLE II
**RANDOMNESS TEST FOR SHOPPING BEHAVIOR DATA**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Quantitative Data</th>
<th>Ordinal &amp; Nominal Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV &amp; DV</td>
<td>Adjacent Test</td>
<td>Ordinal Test: Z ≤ 1.65</td>
</tr>
<tr>
<td>y</td>
<td>1.73</td>
<td>0.91</td>
</tr>
<tr>
<td>x₁</td>
<td>2.17</td>
<td>0.62</td>
</tr>
<tr>
<td>x₂</td>
<td>1.64</td>
<td>1.35</td>
</tr>
<tr>
<td>x₃</td>
<td>1.84</td>
<td>-0.91</td>
</tr>
<tr>
<td>x₄</td>
<td>1.94</td>
<td>-1.97</td>
</tr>
</tbody>
</table>

The third test involves the verification of whether the data contain any extreme values. Since regression modeling is based on ordinary least square (OLS) method, significant outliers in the data array would cause large fluctuation of the values about the mean and, thus, makes the forecast unreliable. Since we are using linear regression modeling in this study, we tested all data sets in IV and DV for extreme values. It was found that there are no extreme values.

For online shopping behavior study, this finding of no extreme value implies that the respondents provided candid opinions to the questionnaires. Secondly, the non-existence of extreme value in the data set also signifies that the data set does not contain any defective response. Extreme values are considered defective response. A survey containing extreme values must be removed from the set and replaced with a new survey. Extreme values reduce the reliability of the data, and may become a source of inferential error in the analysis.

### TABLE III
**EXTREME VALUES AND SIGNIFICANT OUTLIERS**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean Value</th>
<th>Sample STD</th>
<th>Zmax</th>
<th>Zmin</th>
<th>Max/Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>2.05</td>
<td>0.68</td>
<td>1.40</td>
<td>3.02</td>
<td>-0.96 / 1.42</td>
</tr>
<tr>
<td>x₁</td>
<td>1.91</td>
<td>0.61</td>
<td>1.77</td>
<td>-1.49</td>
<td>-0.23 / -5.57</td>
</tr>
<tr>
<td>x₂</td>
<td>2.12</td>
<td>1.00</td>
<td>0.88</td>
<td>-2.11</td>
<td>-1.24 / -4.23</td>
</tr>
<tr>
<td>x₃</td>
<td>1.72</td>
<td>0.64</td>
<td>1.22</td>
<td>-2.69</td>
<td>-0.78 / -6.89</td>
</tr>
<tr>
<td>x₄</td>
<td>1.66</td>
<td>0.66</td>
<td>2.04</td>
<td>-2.52</td>
<td>0.58 / -6.33</td>
</tr>
</tbody>
</table>

*Define extreme value as  Pr(|Z| > 1.65); extreme if Z > 1.65.

A final preliminary test is a test for the central limit theorem. According to CLT, if the data drawn from a random process is adequately large, the data will manifest normal distribution. Thus, CLT help us to verify that the sample used for the study was adequate. If CLT characteristic is manifested, it also confirm our finding that the data came from a random process (Table III) and that the data is normally distributed (Table II). The implication of a finding of CLT implies that the IV and DV data sets are ready for linear regression modeling. What is left is the result of the significance testing of the proposed model.

For online shopping behavior study, this finding of CLT means that people’s opinion about online shopping came from a random process, i.e. different people express different opinions, but ultimately the random process that generated different opinion also allow us to see a picture in a larger context under normal distribution. This finding provides optimum condition for OLS modeling.
TABLE IV
CENTRAL LIMIT THEOREM CONFIRMATION

<table>
<thead>
<tr>
<th>Variables DV &amp; IV</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$Z = \frac{\bar{X} - \mu}{\sigma}$</th>
<th>$\Phi(\frac{Z}{\sigma})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td>2.05</td>
<td>1.89</td>
<td>0.68</td>
<td>2.43 = 0.992</td>
</tr>
<tr>
<td>$X_1$</td>
<td>1.91</td>
<td>1.77</td>
<td>0.61</td>
<td>2.70 = 0.997</td>
</tr>
<tr>
<td>$X_2$</td>
<td>2.12</td>
<td>1.89</td>
<td>1.00</td>
<td>1.65 = 0.950</td>
</tr>
<tr>
<td>$X_3$</td>
<td>1.72</td>
<td>1.57</td>
<td>0.64</td>
<td>2.57 = 0.995</td>
</tr>
<tr>
<td>$X_4$</td>
<td>1.66</td>
<td>1.51</td>
<td>0.65</td>
<td>2.54 = 0.994</td>
</tr>
</tbody>
</table>

*Normal $Z = -1.645$ if $\Phi(z) = 0.05$.

We begin the modeling process with simple regression where each IV is regressed against the DV. It was found that access to the online shopping platform (X1) does not have significant association with online shopping behavior. This finding seems to be counter-intuitive because online shopping platform is the requisite for online shopping. Other IVs: intent to shop online (X2), attitude on online shopping (X3) and perception towards e-commerce infrastructure in Thailand (X4) significantly relate to online shopping behavior.

The finding of significant relationship between X2, X3 and X4 with online shopping behavior (Y) was not surprising. However, the non-significance relationship between X1 and Y requires further examination. This further examination was accomplished in multiple regression. We speculate that if X1 becomes a significant co-determinant in multiple regression modeling, it implies that X1 interacts with other IVs (X2, X3 and X4) to produce Y.

In the second stage of the modeling process, we began to add different combinations of IVs into the modeling with two variables combination and then finally four variables.

It was found that in the proposed model consisting of all four Xs, the significance test under ANOVA, the reading for F test is 2.84 compared to the theoretical value of 1.69. Where under simple regression X1 was insignificantly related to online shopping behavior (Y), but under multiple regression we saw significant relationship of all Xs. The final proposed model for online shopping behavior is:

$$ Y = 0.04 - 0.09X_1 + 0.18X_2 + 0.15X_3 + 0.32X_4 + e $$

However, the individual reading of the statistical significance of each variable forces us to drop X1 and X3 from the model. The final model where each IV is statistically significant is:

$$ Y = 1.04 + 0.20X_2 + 0.36X_4 + e $$

This finding implies that without the proposed explanatory factor, the propensity for online shopping in Thailand is 1.04 (low level in a scale of [0, 1, 2, 3]). Intent and e-commerce infrastructure play a positive influence on online shopping behavior. Intent to shop online contributes 0.20 times and perception toward e-commerce infrastructure contributes 0.36 times to online shopping behavior. The mean value for the scale (0, 1, 2, 3) is 1.5. If we substitute 1.5 for each X, the result is:

$$ Y = 1.04 + 0.20(1.5) + 0.36(1.5) + e $$

or

$$ 1.88 \pm e $$

The error value in this model is $\pm 0.63$. Thus, the forecast based on the mean value for the model is as high as $1.88 + 0.63 = 2.51$ or as low as $1.88 - 0.63 = 1.25$. The TRACK ON INTERNET OF THINGS OF INRIT 2016, Impact Forum, SAPPHIRE 118, BANGKOK THAILAND, 6 July 2016
### TABLE VI
MULTIPLE REGRESSION OF ONLINE SHOPPING BEHAVIOR IN THAILAND

<table>
<thead>
<tr>
<th>Proposed Model</th>
<th>$F(\text{obs})$</th>
<th>$F(\text{crit})$</th>
<th>Conclude</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + e$</td>
<td>0.91</td>
<td>1.69</td>
<td>Not significant</td>
</tr>
<tr>
<td>$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + e$</td>
<td>0.80</td>
<td>1.69</td>
<td>Not significant</td>
</tr>
<tr>
<td>$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + e$</td>
<td>1.22</td>
<td>1.69</td>
<td>Not significant</td>
</tr>
<tr>
<td>$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + e$</td>
<td>1.18</td>
<td>1.69</td>
<td>Not significant</td>
</tr>
<tr>
<td>$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_6X_6 + e$</td>
<td>2.46</td>
<td>1.69</td>
<td>Significant</td>
</tr>
<tr>
<td>$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 + \beta_7X_7 + e$</td>
<td>1.82</td>
<td>1.69</td>
<td>Significant</td>
</tr>
<tr>
<td>$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 + \beta_7X_7 + \beta_8X_8 + e$</td>
<td>1.42</td>
<td>1.69</td>
<td>Not significant</td>
</tr>
<tr>
<td>$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 + \beta_7X_7 + \beta_8X_8 + \beta_9X_9 + e$</td>
<td>2.53</td>
<td>1.69</td>
<td>Significant</td>
</tr>
<tr>
<td>$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 + \beta_7X_7 + \beta_8X_8 + \beta_9X_9 + \beta_{10}X_{10}$</td>
<td>1.83</td>
<td>1.69</td>
<td>Significant</td>
</tr>
<tr>
<td>$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 + \beta_7X_7 + \beta_8X_8 + \beta_9X_9 + \beta_{10}X_{10} + \beta_{11}X_{11} + e$</td>
<td>2.75</td>
<td>1.69</td>
<td>Significant</td>
</tr>
<tr>
<td>$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 + \beta_7X_7 + \beta_8X_8 + \beta_9X_9 + \beta_{10}X_{10} + \beta_{11}X_{11} + \beta_{12}X_{12} + e$</td>
<td>2.84</td>
<td>1.69</td>
<td>Significant</td>
</tr>
</tbody>
</table>

Recall that when $X_1$ was regressed against $Y$, it was found that there was no significant relationship. However, when $X_1$ is included in the multiple regression model, it became a significant explanatory factor for the model. We determine the interaction between $X_1$ and other IVs: $X_2$, $X_3$ and $X_4$. It was found that $X_1$ (accessibility) significantly interacts with $X_2$ (intent to shop online), $X_3$ (attitude towards online shopping) and $X_4$ (perception of e-commerce infrastructure in Thailand).

This finding has a significant implication for online shopping. Accessibility compliments other factors to determine online shopping behavior. Despite the significant interaction between $X_1$ and other IVs: $X_2$, $X_3$ and $X_4$. It was found that $X_1$ (accessibility) significantly interacts with $X_2$ (intent to shop online), $X_3$ (attitude towards online shopping) and $X_4$ (perception of e-commerce infrastructure in Thailand).

The determination of gender and age according to IV factors: accessibility ($X_1$), intent to shop online ($X_2$), attitude towards online shopping ($X_3$) and perception towards e-commerce in Thailand ($X_4$) shows low correlation coefficient and are statistically insignificant. These findings imply that online shopping is a generalized mass market. Segmentation on the basis of gender or age is not appropriate.

The accuracy of each variable was tested for its likelihood estimation. Using the QQ plot method and the value of $eta$ as the indicator for estimated mean, the observed mean is compared with $eta$. It was found that $X_3$ (attitudes towards online shopping) shows statistical significant difference between the observed and estimated values. This finding implies that people’s attitude towards online shopping varies significantly.

Since $X_1$ and $X_3$ had been removed from the final model, the remaining $X_2$ and $X_4$ are the ones that would play a major role in our

### TABLE VII
INTERACTION EFFECT AMONG INDEPENDENT VARIABLES

<table>
<thead>
<tr>
<th>Possible Interaction Effect</th>
<th>$T_{\beta_2}$ (obs)</th>
<th>$T_{\beta_2}$ (crit)</th>
<th>Conclude</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_2X_1X_2$</td>
<td>3.53</td>
<td>1.64</td>
<td>Significant</td>
</tr>
<tr>
<td>$\beta_2X_1X_3$</td>
<td>2.96</td>
<td>1.64</td>
<td>Significant</td>
</tr>
<tr>
<td>$\beta_2X_1X_4$</td>
<td>3.39</td>
<td>1.64</td>
<td>Significant</td>
</tr>
<tr>
<td>$\beta_2X_2X_4$</td>
<td>-1.84</td>
<td>1.64</td>
<td>Significant</td>
</tr>
</tbody>
</table>

From the demographic information we collected, age and gender were used as indicator for online shopper segmentation. It was found that neither age nor gender plays a significant role on online shopping behavior: “gender” correlation coefficient $r_{pb} = -0.03$, $r_{rb} = 0.02$ and $r_{\phi} = 0.10$; and “age” correlation coefficient $r_{pb} = -0.11$, $r_{rb} = -0.01$ and $r_{\phi} = 0.06$. All $T$ test results for gender and age shows no statistical significance.

The accuracy of each variable was tested for its likelihood estimation. Using the QQ plot method and the value of $eta$ as the indicator for estimated mean, the observed mean is compared with $eta$. It was found that $X_3$ (attitudes towards online shopping) shows statistical significant difference between the observed and estimated values. This finding implies that people’s attitude towards online shopping varies significantly.

Since $X_1$ and $X_3$ had been removed from the final model, the remaining $X_2$ and $X_4$ are the ones that would play a major role in our
interpretation of the estimated means and observed mean values:

\[ Y = 1.04 + 0.20X_2 + 0.36X_4 + e \]

These two variables X2 and X4 show no significant difference between the observed and estimated value. This finding implies that the responses from the survey is predictable within acceptable range of accuracy, i.e. 0.95 confidence interval: \( Z = (\eta - \bar{X}) / S \).

### VI. DISCUSSION

This study begins with four proposed explanatory factors: X1 (accessibility to e-commerce platform), X2 (intent to shop online), X3 (attitude towards online shopping), and X4 (perception toward e-commerce infrastructure in Thailand). In the final analysis, only X2 and X4 are statistically significant in determining online shopping in Thailand. The proposed model including the interaction effect term is:

\[ Y = 1.04 + 0.20X_2 + 0.36X_4 - 1.14X_2X_4 + e \]

![Fig 2. Path Diagram for CFA for Online Shopping in Thailand](image)

The intercept of 1.04 implies that there is a low level of online shopping in Thailand. The key for this interpretation comes from the scale definition: 0 = none, 1 = low, 2 = medium, and 3 = high. Variables X2 and X4 positively contribute to online shopping behavior at a factor of 0.20 and 0.36, respectively. This finding implies that stakeholders, such as web store owners should focus on e-commerce infrastructure (X4), i.e. security, privacy, convenience, accessibility, etc.

### VII. CONCLUSION

This research looks at online shopping behavior in the consumer perspective through survey. Thailand is used as a case study. We found that online shopping behavior depends on online shopper’s attitude towards e-commerce. This component of IV is “consumer side” determinant. A second determinant is e-commerce infrastructure; this co-factor is “supplier side” determinant. For stakeholders involving in online shopping, this finding points to the targets on which policy and planning should be aimed. The result of
this research tells stakeholders what online consumers behavior are the drivers for online shopping in Thailand: intent to shop and perception towards e-commerce infrastructure.

REFERENCES

(Arranged in the order of citation in the same fashion as the case of Footnotes.)


