Cost estimation methods for internet infrastructure deployment in Rural Sarawak: a review

Abel Chai Yu Hao, Chean Hung Lai, Hieng Tiong Su, Yi Lung Then, Fei Siang Tay, Patrick Hang Hui Then, Natalie Lim Chen Yi
Faculty of Engineering, Computing and Science, Swinburne University of Technology Sarawak Campus, Malaysia

Article Info
Article history:
Received Jan 20, 2020
Revised Mar 22, 2020
Accepted Apr 12, 2020

Keywords:
Cost estimation methods
Parametric model
Probabilistic cost model
Rural development
Sarawak

1. INTRODUCTION
The Malaysia government promised to improve the internet penetration rate in the country surpasses 90% in 2020 [1]. The leading telecommunication infrastructure company in Sarawak will construct 600 telecommunication towers in 2019 to improve the network penetration in Sarawak [2]. This will be the first phase of 5000 telecommunication towers that Sarawak government intends to build in Sarawak. Sarawak government also inked Memorandum of Understanding (MoU) with a few global companies to strength the digital transformation of Sarawak. With fast improving internet technologies, the digital divide between rural area and urban area is still a critical issue all around the world. With current high-speed internet connection, rural area seems disconnected from our society and unable to enjoy the benefits from high-speed internet.

The main factors that hinders internet connectivity in rural areas are the high installation and maintenance cost, low population density, urbanization, lack of resources and challenges in transportation [3]. With current internet technologies, rural can have potential to obtain some benefits including promoting literacy, improving health care, reducing market inefficiencies, increase government transparency and enabling environmental monitoring. There are some wireless technologies such as Wi-Fi, WiMax, LoraWan and TV White Space. The maximum range of outdoor transmission for Wi-Fi is low and only can be enhanced by using stronger directional antennas. This will indirectly affect the infrastructure cost of the network implementation. The WiMax technology requires high amount of sub-towers to extend the network coverage which result in high implementation cost. LoraWan that utilizes the license-free ISM band for rural connectivity but they have slower data transfer rate and requires stronger line-of-sight link [4]. TV White Space technology utilizes the unused TV band for data transmission. The TV White Space technology has high range of transmission with limited bandwidth.

Corresponding Author:
Abel Chai Yu Hao,
Faculty of Engineering, Computing and Science,
Swinburne University of Technology Sarawak Campus,
Jalan Simpang Tiga, 93350 Kuching, Sarawak, Malaysia.
Email: abel_cyh94@hotmail.com

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.
The rural internet development is not prioritised because of low population density and return of investment from urban area development is more favourable. The general cost estimation methods seldom concentrate on rural characteristics which result in imbalance between network performance and total project cost for ISPs. There are a few cost estimation methods with different characteristics, accuracy and sensitivity. A suitable cost estimation methods will affect the accuracy and reliability of the estimated cost [5-8]. Therefore, the objective of the paper is to review and evaluate the viability and existing cost estimation methods for risk management and economic analysis particularly for rural internet implementation in Sarawak. The rest of the paper is organised as follows; section 2 presents various cost estimation methods, section 3 describes characteristics of rural Sarawak and proposes suitable cost estimation methods.

2. COST ESTIMATION METHODS

The cost estimation methods can be categorized into three aspects which are expertise judgement, parametric equation and mathematical algorithm as shown in Figure 1. Cost estimation methods also can be grouped into algorithmic models and non-algorithmic models. Algorithmic models rely on statistical or numerical analysis of historical data while non-algorithmic models focus on case-by-case analysis.

2.1. Traditional cost estimation

This method requires quotations from various vendors or historical data to obtain the actual cost of each materials and equipment. The total cost of the materials and equipment are defined as initial starting point of estimations for the project. The estimations are adjusted according to amount of materials, labour, equipment and other cost item [9]. This will become the base-line for the total cost of the project [10]. The total cost will be further increase to overcome the inflation of labour, equipment and material. The final cost will be added some contingency amount (normally set as certain percentage depends on allocated risk) to handle uncertain events. Therefore, the estimated cost is the summation of initial starting estimation, inflation rate and contingency amount. This estimation methods usually used as reference point or actual project cost to compare the accuracy of other estimation methods. From this method, the accuracy of the estimation is high since all information collected are valid and latest. Collection of information from possible vendors and historical data are the key factor to determine the duration required and accuracy for this estimation. Normally, every different project requires different materials, labour, equipment and vendors. So, each projects require to collect new information from vendors and historical data [11]. This means that all projects requires to undergo the estimation of initial starting point estimations, inflation rate estimations and contingency estimations. This estimation method unable to provide the probability of over budget or below budget and determine the key factors that most influential in total project cost. Therefore, this estimation method only useful for estimating total cost of the project without further analysing the cost driver.

2.2. Engineering experience

This estimation method usually performed by experienced personal which participated in similar project. Those engineers obtained engineering experiences from previous similar project and deployed in current project [12-14]. The current project cost will be scaled up or down depending on the size of current project to previously deployed project. The engineer required to determine the scaling factors and scaling...
percentages according to their opinions and experiences. After the scaling factors and scaling percentages have been decided, the project cost will be further increased by certain percentage to handle other unforeseen event. The extra cost added is defined as contingency cost from risk management [15]. The cost estimation contingency ensure that the estimated project cost is able to resolve any uncertainties encountered. But, adding percentage as contingency cost may lead to oversight some risk. Since the scaling factors and scaling percentages are determined by the engineers without solid evidence or depends on historical data, the outcome project cost may be over budgeted. This method is able to identify the minor differences between both projects and helps in knowing their impacts [16, 17].This estimation method only applicable to current projects that are similar with previously deployed projects or insufficient information to perform other estimations [13, 18].

2.3. Mean least square regression models

Mean Least Square (MLS) regression models is a statistical method that analyse the relationship between one or more independent variables and the dependent variables [19]. For network implementations, the dependent variables will be the project cost while the independent variables are the cost factors. MLS regression models utilise the historical data. The MLS regression model’s equation is expressed by (1).

\[ Y = a + b_1x_1 + b_2x_2 + \cdots + b_nx_n \]  

where \( Y \) is the dependent variables, \( a \) is the regression constant, \( b_1, b_2, \ldots, b_n \) are regression estimates and \( x_1, x_2, \ldots, x_n \) are independent variables. The regression constant indicates the intersection point between \( Y \)-axis and \( X \)-axis. The regression estimates indicates the degree of steepness of the independent variables.

The MLS regression model is suitable for small number and grouped data. When the number of data is large, this model tends to underestimates the larger data and overestimates the smaller data as MLS regression model regresses towards the mean value [20]. The project cost for network implementations are always scatted and depends on the design. MLS regression model performed estimations with assumptions that the data collected is normally distributed. The MLS regression model is also sensitive to data outlier. MLS regression model is more accurate when the data size is small. So, data that further from the average value are defined as outliers. There are a few methods to determine the data outliers for example Row Deletion Methods and Forward Search [21]. Therefore, a single outlier can changed the regression estimates and coefficient of determination, \( R^2 \) of the MLS regression model [20].

2.4. Parametric model

Parametric model performs cost estimations with mathematical algorithms. All materials and components are analysed and formulated into Cost Estimating Relationship (CER). The CERs are mathematical representation of the relationship between cost drivers and total cost [13, 14, 22, 23]. The complete development of a parametric model undergoes scope determination, data collection, data normalization, data analysis, data application, testing and documentation [24]. The scope is important to define the functions, characteristics and cost drivers of the project [25]. The low level data collected must be normalised before further analysing and to be used for the parametric model. For example, all the cost must be at the same base year. This can increase the accuracy of the model developed. The data analysis consist of various methods to determine the coefficient of determination, \( R^2 \) of each cost drivers to the project cost [26]. \( R^2 \) that is higher than 0.9 is acceptable while value between 0.6 to 0.9 requires further analysis [27]. This steps is usually can be a time-consuming process. This coefficient of determination, \( R^2 \) is a measurement to show how accurate the equation represent the variability of the input data.

2.5. Capacity factored estimate model

CFE model is able to estimate new project cost from previously similar project when the capacity for both projects are known. This CFE highly relies on the non-linear relationship between project capacity and project cost. The CFE model’s equation is expressed by (2).

\[ C_{\text{new}} = C_{\text{previous}} \times \left( \frac{\text{Cap}_{\text{new}}}{\text{Cap}_{\text{previous}}} \right)^e \]  

where \( C_{\text{new}} \) is the estimated cost of the new project, \( C_{\text{previous}} \) is the actual cost of the previous project, \( \text{Cap}_{\text{new}} \) is the capacity of the new project \( \text{Cap}_{\text{previous}} \) is the capacity of the previous project and \( e \) is the scaling factor.
The scaling factor normally lies between 0.5 and 0.85 [28]. The scaling factor of CFE is defined as the slope of the cost curve that represents the change in cost when the new project is larger or smaller than the previous project. The main challenge in using CFE method is selecting correct scaling factors. The previously similar project must be close in size and capacity to the new project. Some external factors such as location, environment and time will further reduce the accuracy of this CFE method. In order to increase the accuracy of the estimations, cost index is used to overcome the inflation rate when the previous project was completed few years before. The suitable scaling factor can be determined by performing stimulation with large number of previously known project. This CFE method can be prioritise when only limited information available. But, this may result in negative impact when the neglected parameters are very important in cost estimations [29, 30].

2.6. Neural network

Neural network model also able to solve regression problem. Neural network is a network or circuit that can learn to solve artificial intelligence (AI) or regression problems. Among all different neural network, multilayer feedforward neural networks are the most popular method to handle regression problems. Multilayer feedforward networks feed the input into hidden units and the output from hidden units will eventually feed to the output units. The Figure 2 shows a multilayer feedforward neural network with one hidden layer [31].

![Figure 2. Multilayer feedforward neural network with one hidden layer](image)

From Baum [32], the hidden units and weights can be added to the training algorithm of the neural network to enable the neural network solves any other learning problems. Therefore, with suitable hidden units and weights applied, the neural networks become universal learner that able to solve any solvable problems. There are different algorithms to alter the neural network architecture to improve the learning process for example pruning algorithms [33] and constructive algorithms [34]. The constructive algorithms able to search for small network solutions first and less training time required than pruning algorithms. The accuracy and performance of neural networks are highly depending on the data collected. If the data collected is not complete, the neural network unable to generate suitable algorithms to train the model for cost estimations. Therefore, neural network requires big data to produce an accurate model for estimations and this requires long processing time. In addition, the selection of suitable training algorithms such as numbers of hidden units is a major challenge for this neural network.

2.7. Probabilistic cost model (PCM)

Most of the cost estimation model only concentrate on cost drivers that directly contribute in total project cost. Unexpected events are always ignored during the estimation process and this is actually important when preforming real cost estimation with high accuracy [14, 35]. PCM model is used to estimate the total cost for unexpected events during the project lifetime. Unexpected events such as expansion site or material price change will induce some extra cost to total project cost. Those unexpected events are normally unpredictable and random in nature which increase the challenge for all cost model to estimate an accurate total project cost. Previous cost estimation methods handle the extra cost from unexpected events by including some contingency cost in total project cost. This is the easiest and fastest method to handle the extra cost but may result in overestimates or underestimates of total project cost and the unexpected events are not clearly identified. After identified all possible unexpected events, PCM model is able to identify the total cost for unexpected events.
For PCM model, the unexpected events are assumed to be independent from each other and only one event occurred at a time. This enable the unexpected events fit in Poisson distribution and modelled into Poisson process. The probability of having cost overrun can be determined when the contingency, β is set. The probability can be estimated by using the (3).

\[
P(C_t \leq \beta) \geq p, \beta = 1 - p
\]

\[
P(C_t \leq \beta) = \sum_{x=0}^{\infty} P(C_t \leq \beta | X = x) P(X = x)
\]

\[
P(C_t \leq \beta) = \sum_{x=0}^{\infty} \Phi \left( \frac{\beta - x \mu}{\sqrt{\mu x}} \right) \frac{\lambda^x e^{-\lambda}}{x!} \geq p
\]

where \(C_t\) is the total cost of unexpected events, \(\mu_c\) is the average cost of unexpected events, \(\lambda\) is average rate of event occurred, \(\Phi\) is cumulative distribution function of \(x\), \(C\) is the initial total cost of project.

The probability of having cost overrun can be determined when the contingency, \(\beta\) is set. The probability can be estimated by using the (3).

\[
P(X \leq x) = \sum_{i=0}^{\infty} \Phi \left( \frac{\beta - x \mu}{\sqrt{\mu x}} \right) \frac{\lambda^x e^{-\lambda}}{x!} \geq p
\]

where \(C_t\) is the total cost of unexpected events, \(\mu_c\) is the average cost of unexpected events, \(\lambda\) is average rate of event occurred, \(\Phi\) is cumulative distribution function of \(x\), \(C\) is the initial total cost of project.

The estimated cost of unexpected events can be estimate since the unexpected events follows a Poisson distribution with \(\lambda\). The initial cost estimation of unexpected events can be obtained with (4).

\[
E(C_t) = E(\sum_{i=1}^{\infty} C_i)
\]

\[
E(C_t) = E(X)E(C_i)
\]

3. DISCUSSION AND ANALYSIS

3.1. Characteristics of Rural Sarawak

The population of Sarawak increased constantly from 2.70 million in 2015 until 2.81 million in 2019 [36]. Sarawak is the largest state in Malaysia which covering an area of 124,449.51 square kilometres. Sarawak’s area is almost equal to Peninsular Malaysia. The large area of Sarawak results in poor Internet penetration and slower digital transformation. Researchers from Swinburne University of Technology, Melbourne, Australia and Swinburne University of Technology Sarawak Campus had presented a research report of digital inclusion and mobile media in rural Sarawak. The research took place in 11 rural villages in Sarawak. The report highlighted that only 36 per cent of residents in rural Sarawak have accessed to the Internet compared with 56 per cent from overall Malaysia [37]. This shows that our Sarawak are still lagging behind from Internet penetration and digital transformation. Our Government provided a few benefits or welfare schemes to those in need which included child benefits, aged benefits, handicapped benefits and poverty assistance [37]. The research discovered that at least 61 per cent of resident from rural area received one of the support from the government. Furthermore, only 18 per cent of residents from rural area did not own any devices that can access to the Internet [37]. The characteristics of Rural Sarawak as shown in Table 1.

<table>
<thead>
<tr>
<th>Characteristics of Rural Sarawak</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of populations in 2019 (million)</td>
<td>2.81</td>
</tr>
<tr>
<td>Area (km²)</td>
<td>124,449.51</td>
</tr>
<tr>
<td>Rate of rural resident accessed to the internet (%)</td>
<td>36</td>
</tr>
<tr>
<td>Rate of rural resident received support of welfare schemes from Government (%)</td>
<td>61</td>
</tr>
<tr>
<td>Rate of rural resident did not own any Internet support device (%)</td>
<td>18</td>
</tr>
</tbody>
</table>

The development of rural area Sarawak can be determined from a few aspects which are electricity, water, supply, schools, health clinics, road connection and mobile coverage. Older residents treat electricity, water, supply, schools and health clinics as the key elements in rural development. Rural connectivity is not top priority among older residents but the younger generations claim that rural connectivity is also one of the important aspect and should be emphasized [38]. Generally, road connection is always used as a parameter to measure how “developed” an area is. Connectivity has become a hot issue among countries and some countries start to measure their rural development according to rural connectivity. Sarawak tends to reduce the rural-urban development gap by launching the Sarawak Digital Economy Strategy 2018-2022 which includes 47 plans to drive Sarawak into high-income state in Malaysia.
The residents in rural Sarawak normally have their own living style to make a living such as agriculture, mining, timber sectors, fishing industry and local tourism. Most of the rural residents have met the minimum income standard but their living environment can be further improved by our current technologies. By promoting internet to rural areas, they are able to receive the latest information from time to time which prevent them feel like excluded from outside world. They can gain new knowledge or methods to improve their current lifestyles. The government can establish stronger connection with rural areas while seek for their needs. The education and health issue also can be improved with enhanced rural connectivity. Sarawak is the largest state in Malaysia which result in most of the rural areas are far from urban cities. Therefore, implementation of Internet connectivity requires strong backhauling system which indirectly increases the total project cost. The high initial investment and unforeseen return of investment reduce the interest of Internet Service Provider (ISP) to invest in improving rural connectivity for rural Sarawak.

3.2. Proposed cost estimation methods for Rural Sarawak

The government promised to improve the internet penetration rate in the country surpasses 90% in 2020 [4]. The rural areas have higher potential in increasing internet penetration rate since internet penetration rate in urban areas are consider to be high and saturated. The low-cost internet will induces momentum to the number of users of internet in rural areas and increase the number of internet users. This will gradually attracts more internet service providers to invest in rural areas. The competition between ISPs will further reduce the cost of internet which benefits the residents in rural areas. Comparison between different cost estimation methods as shown in Table 2.

<table>
<thead>
<tr>
<th>Cost estimation methods</th>
<th>Time required</th>
<th>Information required</th>
<th>Weakness</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional cost estimation</td>
<td>High</td>
<td>High</td>
<td>Must have all information for high accuracy, Case by case basic, Lack of further analyses of cost driver</td>
<td>High</td>
</tr>
<tr>
<td>Engineering experience</td>
<td>Low</td>
<td>High</td>
<td>Influence by personal instinct, Case by case basic, Only applicable when current project is similar with previous project</td>
<td>Low</td>
</tr>
<tr>
<td>Mean Least Square regression model (MLS)</td>
<td>moderate</td>
<td>moderate</td>
<td>Only suitable for small number and grouped data, Tends to underestimate the large data and overestimates the smaller data, Sensitive to data outlier</td>
<td>moderate</td>
</tr>
<tr>
<td>Parametric model</td>
<td>High</td>
<td>High</td>
<td>Requires all cost driver information, Data normalisation and data analyse can be time-consuming process.</td>
<td>High</td>
</tr>
<tr>
<td>Capacity Factored estimation model (CFE)</td>
<td>moderate</td>
<td>Low</td>
<td>External factors such as environment and time will affect the accuracy, Only prioritise when limited information available, Reduced accuracy when neglected or missing parameters are very important, Difficult to design suitable training algorithm, Requires big data</td>
<td>Low</td>
</tr>
<tr>
<td>Neural network</td>
<td>High</td>
<td>Moderate</td>
<td>Accuracy varies with data collected, Estimation only include unexpected events during the project lifetime</td>
<td>High</td>
</tr>
<tr>
<td>Probabilistic Cost Model (PCM)</td>
<td>High</td>
<td>High</td>
<td></td>
<td>High</td>
</tr>
</tbody>
</table>

The accuracy of cost estimation methods are subjective to the availability of information of low level cost driver. The traditional cost estimation method and engineering estimation method are case-by-case basis and requires all information starting from material used until design topologies to produce an accurate total project cost estimation. These methods requires longer time and some projects unable to provide all information at the initial stage especially when the project carried out by collaboration of a few companies. The Mean Least Square regression model only effective where the data is small and similar. From previous sections, Sarawak has large area which means all rural areas have different characteristics and distance to urban areas. Therefore, the data will be scattered and not suitable for Mean Least Square regression model. CFE model is suitable for total project cost estimation when not all low level cost drivers are available.
This method can be used as initial estimation methods to determine the feasibility of the project before spending time and resources to determine all low level cost drivers. After collecting all low level cost drivers, parametric model can be used as detailed estimation to obtain a more accurate estimated total project cost. PCM can be used to determine the total cost for unexpected events to prevent underestimation or overestimation of the total project cost.

4. CONCLUSION

One of the contributions of this paper was to identify and analyse some cost estimation methods for wireless network implementation. This paper also highlighted the characteristics and cost estimation methods suitable for rural Sarawak. The accuracy of various cost estimation methods are depend on the data collected and level of confidence. CFE model is suitable for some projects where not all data is easily available. Parametric model is highly recommended for some projects that requires high accuracy and all low level cost drivers are available. PCM model can be used to estimate the total cost for unexpected events. Sarawak has large area which means all rural areas has different characteristics and distance to urban areas. Therefore, the data will be scattered and only CFE model and parametric model are suitable for total project cost estimation.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Sarawak Multimedia Authority (SMA) for financing this research through Sarawak Digital Economy Research Grant.

REFERENCES


BIographies of Authors

Abel Chai Yu Hao received his Bachelor of Engineering (Honours) in Electrical and Electronics from Swinburne University of Technology Sarawak Campus in 2018. He is currently a master student from Swinburne University of Technology Sarawak Campus in Master of Engineering (Research). His research interests include wireless technologies, rural connectivity and cost-performance analysis of various wireless technologies.

Chean Hung obtained his Bachelor of Engineering (Honours) in Robotic and Mechatronics Engineering and Ph.D in Engineering from Swinburne University of Technology Sarawak Campus. He is currently a senior lecturer and Discipline Leader for R&M engineering in the Faculty of Engineering, Computing and Science at Swinburne University of Technology Sarawak Campus. His research interests include energy management system and control, hybrid energy storage system, control systems, MEMS, robotics, internet of things and smart systems.
Dr. Su received his BEng degree in Electrical and Electronic Engineering from the University of Liverpool and a PhD degree from the University of Birmingham, UK. After his PhD, he worked as a Research Fellow in the Department of Electronic, Electrical, and Computer Engineering, University of Birmingham. He is currently the Dean of the Faculty of Engineering, Computing and Science. His research interests include wireless communication and the design of rf and microwave passive devices such as compact antenna and filters.

Dr. Then Yi Lung was born in Kuching, Sarawak, Malaysia in 1988. He is a registered chartered engineer with PhD in electrical engineering. Research areas involve instrumentation of electromagnetic wave propagation at microwave frequencies, development of microwave sensors for agricultural applications, renewable energy, and energy-harvesting.

Fei Siang Tay received his BEng (Hons) degree in Electrical and Computer Systems Engineering from Monash University, Malaysia, in 2008, and the PhD degree from the Swinburne University of Technology, Australia, in 2013. His research interests are in the areas of intelligent control systems, smart farming, rural connectivity, dynamical fuzzy modeling of complex system, nonlinear system & control, robotics and dynamical modelling.

Professor Patrick Then the Director for the Centre for Digital Futures Swinburne Sarawak. He has been leading multiple industry-funded projects in research and development in collaboration with prominent ICT partners such as Sarawak Information Systems Sdn Bhd (SAINS), IDS (Malaysia) Sdn Bhd in Sarawak, and organizations around the world. Patrick is a strong advocate of R&D and commercialization of innovations in Big Data, Data Mining and Internet of Things.

Natalie Lim Chen Yi received her Bachelor of Engineering (Honours) in Electrical and Electronics from Swinburne University of Technology Sarawak Campus in 2018. She is currently a master student from Swinburne University of Technology Sarawak Campus in Master of Engineering (Research). Her research interests includes wireless technologies, rural connectivity and path loss modeling for wireless communications.