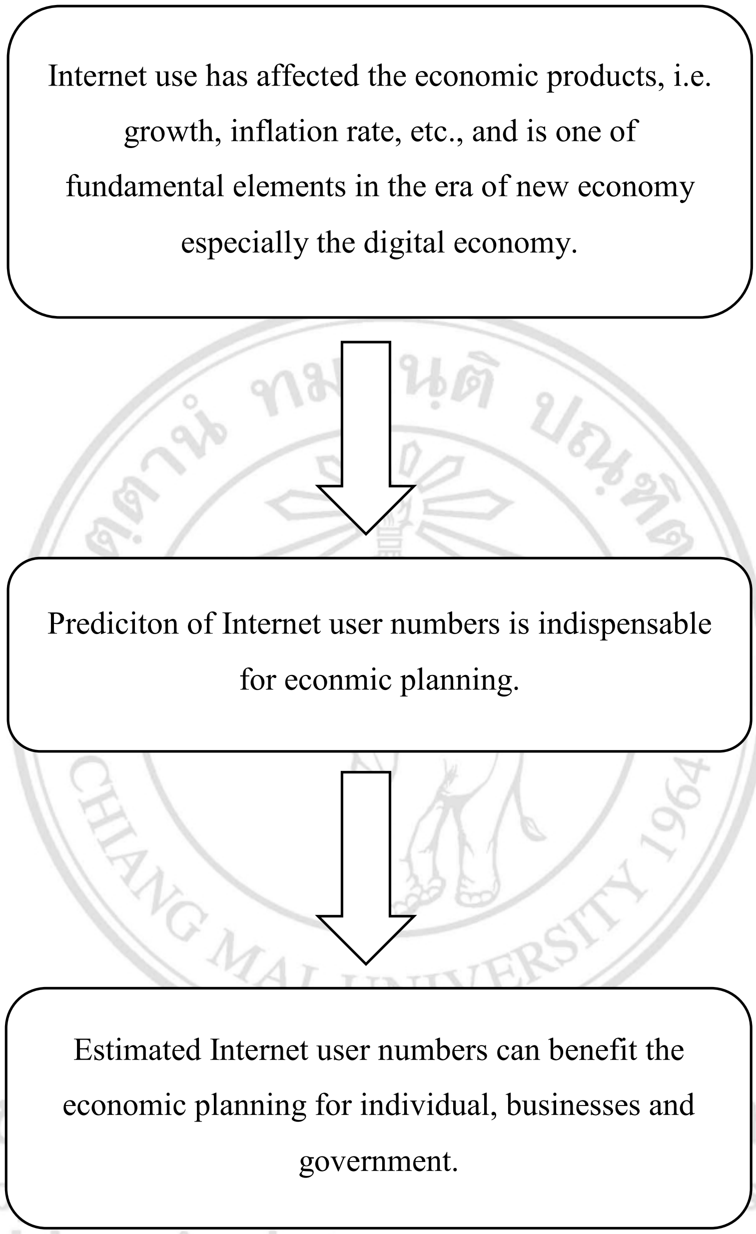
# CHAPTER 3

# Methodology

## 3.1 Conceptual Framework

The study was conducted to determine the complete Stochastic Exponential Growth model for the numbers of Internet users and consequently predict the Internet user numbers in the upcoming years. The data used in the computation were secondary data. Visually, the conceptual framework is given in Figure 2.

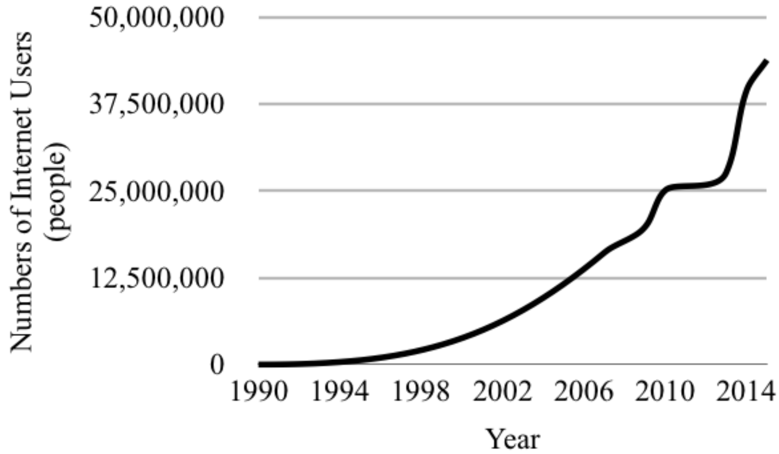



Source: Self-illustration

**Figure 3.1**: Conceptual Framework

### 3.2 Data Collection

The research data in this thesis is the secondary data which were gathered via the National Electronics and Computer Technology Centre (NECTEC) of Thailand. Furthermore, the data were yearly collected from 1991 to 2015. On the contrary, there is the limitation of the data used in this study: the numbers of the data are small. Then, this limitation leads to the use of a Bayesian method which has no assumptions of using a great number of data numbers: in other words, it is appropriate for the study having small sets of data.




Source: NECTEC, 2016

**Figure 3.2**: Numbers of Internet Users in Thailand from 1990 to 2015

**3.3 Research Methodologies/Data Calculating Models**

To start with, the exponential growth model, which is the simplest and the indubitably central model for the population growth research, is proposed. Due to the uncertainties, it is more rational that the growth rate is considered uncertain and that leads to the application of a stochastic process which takes the environmental effects into account. Accordingly, a stochastic exponential growth model is proposed in this thesis, in which the growth rate is in terms of Brownian motion.

The core model of this study is

$$U_t = U_0 \exp\left(\mu t - \frac{\sigma^2}{2} t + \sigma B_t\right) \tag{3.1}$$

where

$U_t$ is the number of Internet users at time $t$.

$U_0$ is the initial number at time $t = 0$.

$\mu$ is the rate of change of the number of Internet users.

$\sigma$ is the volatility.

$B_t$ is one-dimensional Brownian motion.

There are two main parts of this session: the model identification and the prediction using the Bayesian method and Monte Carlo method, including the path generation, respectively.

### 3.3.1 Bayesian Method

The Bayesian method is proposed to estimate the parameters, $\mu$ and $\sigma$ in model (13). Bayesian inference can be obtained in the form of the posterior distribution, which is of the multiplication of likelihood function to the prior (Bishop, 2009), from the Bayes theorem (Harnpornchai & Autchariyapanitkul, 2016).

The posterior distribution's form is

$$f(\mu, \sigma | D) \propto L(D | \mu, \sigma)\, f(\mu, \sigma) \tag{3.2}$$

where

$f(\mu, \sigma | D)$ is the posterior distribution.

$L(D | \mu, \sigma)$ is the likelihood function.

$f(\mu, \sigma)$ is the prior.

D is the research data.



The structure of D is

$$D = (t_1, Y_{t_1}), \dots, (t_N, Y_{t_N}) \tag{3.3}$$

The likelihood function takes the structure:

$$L(D | \mu, \sigma) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \exp[-Z^T \Sigma^{-1} Z] \tag{3.4}$$

where

N is the numbers of data used in the inference.

and

$$Z_k = \frac{1}{\sigma}\left(Y_{t_k} - \mu t_k - \frac{\sigma^2}{2} t_k\right); k = 1, \ldots, N \quad (3.5)$$

and

$$\Sigma_{ij} = E\left[B_{t_i} B_{t_j}\right] = \frac{1}{2}\left(t_i + t_j - |t_i - t_j|\right); i, j = 1, \ldots, N \quad (3.6)$$

**3.3.2 Monte Carlo Method**

According to the inferred parameter using the Bayesian method, the numbers of Internet users at time $t$ can be probabilistically obtained, e.g. $P[N(t) \geq N_{Internet}]$. The probability of such an event is determined using the Monte Carlo method.

Monte Carlo simulation is a method which is applied to simulate a probability space by taking independent samples from the space regarding the probability distribution. (Lee, 2013a) In other words, essentially, the Monte Carlo method requires the generation of random numbers or path in itself.

The Markov Chain Monte Carlo (MCMC) method is conducted to determine the posterior (3.2) (Harnpornchai & Autchariyapanitkul, 2016).

Given (Brandimarte, 2014)

$\theta = [\theta_1 \ \ldots \ \theta_p]$ is a vector of the parameters $\theta_j$ ; $j = 1, \ldots, p$

where

$p$ is the total number of the parameters to be calculated.

The method runs as follow

1. Draw an individual $\theta_{l+1}$ from $q(\theta_{l+1}|\theta_l ; l = 1, \ldots, M)$.

2. Accept the individual $\theta_{l+1}$ with the probability of $\alpha(\theta_l, \theta_{l+1})$; otherwise set $\theta_{l+1} = \theta_l$.

Therefore, the probability of acceptance is

$$\alpha(\theta_l, \theta_{l+1}) = \min\left(\frac{\pi(\theta_{l+1})/q(\theta_{l+1}|\theta_l)}{\pi(\theta_l)/q(\theta_l|\theta_{l+1})}, l\right) \tag{3.7}$$

The independent density, regard a PDF, the next individual does not depend on the current one: $q(\theta_{l+1}) = q(\theta_{l+1})$.

According to the independent density mentioned above, the probability of acceptance is

1. Draw an individual $\theta_{l+1}$ from $q(\theta_{l+1})$.

2. Accept the individual $\theta_{l+1}$ with the probability of $\alpha(\theta_l, \theta_{l+1})$.

Again, the probability of acceptance is

$$\alpha(\theta_l, \theta_{l+1}) = \min\left(\frac{\pi(\theta_{l+1})/q(\theta_{l+1})}{\pi(\theta_l)/q(\theta_l)}, l\right) \tag{3.8}$$