# CHAPTER III

# RESEARCH METHODOLOGY

This chapter presents the research method that consists of several sub chapters i.e research object definition, hidden markov model, data requirement, collecting data method, tools analysis and research framework.

## 3.1 Research Object

The object of this research is prediction on certificated Indonesian Gold price published by Logam Mulia of PT Aneka Tambang, Tbk per gram. Gold price and its causative factors is the topic of this research.

## 3.2 The Hidden Markov Model

### 3.2.1 Hidden Markov Model

Markov chain property is denoted as probability of each subsequent state depends only on what was the previous state:

$$P\ (S_{ik}|S_{i1},S_{i2},\ldots,S_{ik-1}) = P\ (S_{ik}|\ S_{ik-1})$$

States are not visible, but each state randomly generates one of $M$ observations. Based on the figure 3.1 we may conclude that:

a. 'Low' and 'High' atmospheric pressure will be the states, while the observations is 'Rain' and 'Dry'.

b. Transition probabilities from $P$('Low'|'Low') = 0.3 ; $P$('High'|'Low') = 0.7 ; $P$('Low'|'High') = 0.2 and $P$('High'|'High') = 0.8

c. Observation probabilities from $P$('Rain'|'Low') = 0.6 ; $P$('Dry'|'Low') = 0.4 ; $P$('Rain'|'High') = 0.4 and $P$('Dry'|'High') = 0.6

d. Initial probabilities say $P$('Low') = 0.4 , $P$('High') = 0.6

Suppose we want to calculate a probability of a sequence of observations in our example, {'Dry','Rain'}. Consider all possible hidden state sequences:

$P$({'Dry','Rain'}) = $P$({'Dry','Rain'} , {'Low','Low'}) + $P$({'Dry','Rain'} , {'Low','High'}) + $P$({'Dry','Rain'} , {'High','Low'}) + $P$({'Dry','Rain'} , {'High','High'})

Where first term is:

$P$({'Dry','Rain'} , {'Low','Low'})

= $P$({'Dry','Rain'} | {'Low','Low'}) $P$({'Low','Low'})

= $P$('Low') $P$('Dry'|'Low')$P$('Rain'|'Low') $P$('Low'|'Low)

= 0.4*0.4*0.6*0.3

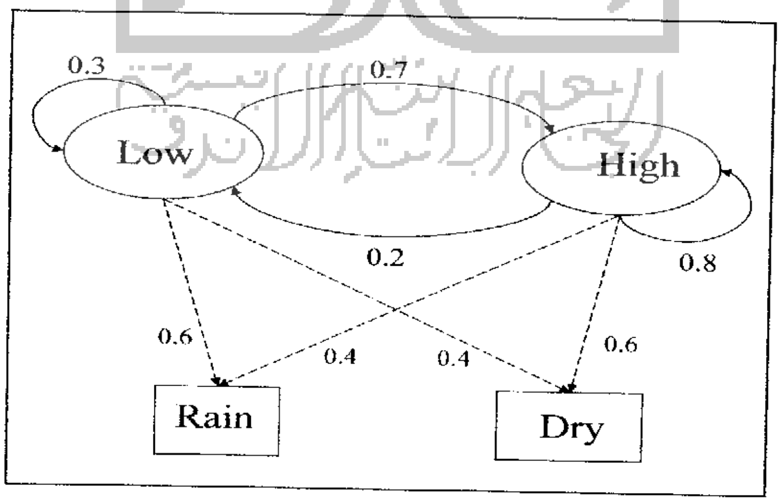


Figure 3.1 Hidden Markov Model

Robert Elliot (1995) proposed three model of Hidden Markov Model:

1. Discrete state and observation (Hidden markov model with discrete time and discrete observation)
2. Continuous range observation (Hidden markov model with discrete time and continuous observation)
3. Continuous-range state and observation (Hidden markov model with continuous time and continuous observation)

The process of observation is denoted with $Y_k$ and causes of events (state) is denoted with $X_k$. Some factors that causing shift in gold rates are dollar exchange rate, political situation, government policies and etc. These events are repeatable however the time is not predictable. In each state, every changes is generated by the random variables $Y_k$ that spread in a certain distribution at interval changes $(\Omega, \mathcal{F}, P)$. State equations and observation equations for the second model is denoted below:

$$X_{k+1} = AX_k + V_{k+1}$$

$$Y_{k+1} = C(X_k) + \sigma(_{k+1})\ \omega_{k+1} \qquad (3.1)$$

Where $\{\omega_{k+1}\}$ is a sequence of stochastic random variables spread identical to normal with average zero and variance one $N$ (0,1). Because $X_k \in S_x$, then the function $C$ and $\sigma$ define by vector $C = (C_1, C_2, \dots, C_n)^T$ and $\sigma\ (\sigma_1, \sigma_2, \dots, \sigma_n)^T$ also $C(X_k) = \langle C, X_k \rangle$ and $\sigma(X_k) = \langle \sigma, X_k \rangle$ with $\sigma_i > 0$, for $1 \le i \le$ N.

In this research, second model from Hidden markov which is discrete state and discrete observation involving the conditional expectation and recursive estimation.

A. Conditional Expectation

Elliot raised several issues related with the conditional expectation:

1. If $X_k = E\,[X_k \mid Y_k]$ is the conditional expectation of $X_k$ if $Y_k$ known, and $\emptyset_i$ (x) $= (2\pi\sigma i)^{-1/2} \exp(-x^2/2\sigma i^2)$ is the probability density function of $N\,(0, \sigma^i)$ for $t \in R$, then there is a conditional distribution

$$P\,(Y_{k+1} \leq t \mid Y_k) = \sum_{i=1}^{N} \langle X_k, e_i \rangle \int_{-\infty}^{-ci} \emptyset i(x)dx \quad \text{(3.2)}$$

2. Conditional density function of $Y_{k+1}$ if given $Y_k$ is

$$\sum_{j=1}^{N} \langle X_k, e_i \rangle\, \emptyset\, j(t-cj) \quad \text{(3.3)}$$

3. While the mutual distribution is

$$P(X_k=e_i, Y_{k+1} \leq t \mid Y_k) = \langle X_k, e_i \rangle \int_{-\infty}^{-ci} \emptyset i(x)dx \quad \text{(3.4)}$$

4. Conditional expectation of $X_k$ if $Y_{k+1}$ known is

$$E\,[X_k \mid Y_{k+1}] = \frac{\sum_{i=1}^{N} \langle X_k, e_i \rangle \emptyset_{i\,(Y_{k+1}-C_i)} e_i}{\sum_{j=1}^{N} \langle X_k, e_j \rangle \emptyset_{j\,(Y_{k+1}-C_j)}} \quad \text{(3.5)}$$

5. Conditional expectation of $X_{k+1}$ if $Y_{k+1}$ known is

$$E\,[X_{k+1} \mid Y_{k+1}] = \frac{\sum_{i=1}^{N} \langle X_k, e_i \rangle \emptyset_{i\,(Y_{k+1}-C_i)} A e_i}{\sum_{j=1}^{N} \langle X_k, e_j \rangle \emptyset_{j\,(Y_{k+1}-C_j)}} \quad \text{(3.6)}$$

B. Recursive Estimation

Recursive estimation is required to estimate the new parameters, which includes estimators for the state, the number of leap, duration of the event and observation process. Each estimator is defined as follows:

1. State estimators $\gamma_{k+1}(X_{k+1}) = \sum_{i=1}^{N} \langle \gamma_k(X_k), \Gamma^i(Y_{k+1}) \rangle\, a_i$ ......... (3.7)
2. Number of leap estimator

   A markov chain moved out of state $e_r$ at time $k$ to state $e_r$ at time $k+1$ with $1 \leq r, s \leq N$ $\langle X_k, e_r \rangle \langle X_{k+1}, e_s \rangle = 1$. For instance $J_{k+1}^{rs}$ is the number of leap from

$e_r$ to $e_s$ until time to $k+1$, then the number of leap will be defined as follow:

$$\gamma_{k+1,k+1}(J_{k+1}^{rs}) = \sum_{i=1}^{N} \langle \gamma_{k,k}(J_k^{rs}), \Gamma^i(Y_{k+1}) \rangle \alpha_i + \langle \gamma_k(X_k), \Gamma^r(Y_{k+1}) \rangle \alpha_{sr} e_s \quad (3.8)$$

3. The duration of the event estimator

For instance $O_k^r$ is the time duration until time $k$, and $X$ is on state $e_r$, $(O_{k+1}^r) = \sum_{n=1}^{k+1} \langle X_{n-1}, e_r \rangle = O_k^r + \langle X_k, e_r \rangle$, then the duration of event will be defined as follow:

$$\gamma_{k+1,k+1}(O_{k+1}^r) = \sum_{i=1}^{N} \langle \gamma_{k,k}(O_k^r), \Gamma^i(Y_{k+1}) \rangle a_i + \langle \gamma_k(X_k), \Gamma^r(Y_{k+1}) \rangle a_r \quad (3.9)$$

4. Observation process estimator

The estimator for observation process will be defined as follow:

$$\gamma_{k+1,k+1}(T_{k+1}^r(Y)) = \sum_{i=1}^{N} \langle \gamma_{k,k}(T_k^r(Y)), \Gamma^i(Y_{k+1}) \rangle a_i + \langle \gamma_k(X_k), \Gamma^r(Y_{k+1}) \rangle Y_{k+1} a_r \quad (3.10)$$

From the model the algorithm will be written as follow:

<u>Step 1.</u>

Set $N$ = cause of event factor, $T$ = observation length and data input $\{Y_k\}$ as same as data table

<u>Step 2.</u>

Determine the initial value

$\pi = (\pi_i)_{2x1}$

$$A = (aji)_{2x2} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

$$C = (Ci)_{2x1} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

$\sigma_I = (\sigma_i)_{2x1} = \binom{\sigma_1}{\sigma_2}$ with $\pi$ = E [Xo] that fulfill $A_\pi = \pi$ and $\sum_{i=1}^{N} \pi i = 1$

Step 3.

Do for $\ell = 1$ until $T$= observation length

a. Set $\alpha_i = Ae_i$, where $e_i$ unit vector at $\mathbb{R}^N$ ........................ (3.11)

$\gamma_0 = (X\text{o}) = \pi$

$\gamma_0 = (J_0^{rs}) = 0$

$\gamma_0 = (O_0^r) = 0$

$\gamma_0 = (T_0^r(y)) = 0$

$\gamma_0 = (T_0^r(y^2)) = 0$

b. Do for $k$=0 until $\ell = 1$

i. Calculate recursive estimators

$\gamma_{k+1}(X_{k+1}) = \sum_{i=1}^{N} \langle \gamma_k(X_k), \Gamma^i(Y_{k+1})\rangle \alpha_i$ ........................ (3.12)

$\gamma_{k+1,\,k+1}(J_{k+1}^{rs}) = \sum_{i=1}^{N} \langle \gamma_{k,\,k}(J_k^{rs}), \Gamma^i(Y_{k+1})\rangle \alpha_i + \langle \gamma_k(X_k), \Gamma^r(Y_{k+1})\rangle \alpha_{sr}$

es ........................................................................................ (3.13)

$\gamma_{k+1}(J_{k+1}^{rs}) = \langle \gamma_{k+1,\,k+1}(J_{k+1}^{rs}), \underline{1} \rangle$ ............................ (3.14)

$\gamma_{k+1,\,k+1}(O_{k+1}^{r}) = \sum_{i=1}^{N} \langle \gamma_{k,\,k}(O_k^r), \Gamma^i(Y_{k+1})\rangle \alpha_i + \langle \gamma_k(X_k), \Gamma^r(Y_{k+1})\rangle \alpha_r$

........................................................................................ (3.15)

$\gamma_{k+1}(O_{k+1}^{r}) = \langle \gamma_{k+1,\,k+1}(O_{k+1}^{r}), \underline{1} \rangle$ ........................ (3.16)

$\gamma_{k+1,\,k+1}(T_{k+1}^{r}(Y)) = \sum_{i=1}^{N} \langle \gamma_{k,\,k}(T_k^r(Y)), \Gamma^i(Y_{k+1})\rangle \alpha_i + \langle \gamma_k(X_k), \Gamma^r(Y_{k+1})\rangle$

$Y_{k+1}\alpha_r$ .................................... (3.17)

$\gamma_{k+1}(T_{k+1}^{r}(Y)) = \langle \gamma_{k+1,\,k+1}(T_{k+1}^{r}(Y)), \underline{1} \rangle$ ........................ (3.18)

with $\Gamma^{(\cdot)}(Y_k) = \dfrac{\emptyset\,\frac{Yk - C(\cdot)}{\sigma(\cdot)}}{\sigma(\cdot)\emptyset(Yk)}\, e(\cdot)$ .................................... (3.19)

$\emptyset(\cdot)$ is the probability density function $N(0,1)$

$\gamma_{k+1}(T_{k+1}\, T_{k+1}) = \gamma_{k+1,\, k+1}(T_{k+1})$ .................................. (3.20)

$\gamma_k(T_k) \quad = \langle \gamma_k(T_k X_k), \underline{1} \rangle$ with $\underline{1} = (1,1,\dots,1)^T \in \mathbb{R}^N$...... (3.21)

ii. Calculate parameter estimators

$$\hat{a}_{sr\,(k+1)} = \frac{\gamma k+1\,(\,)^{rs}_{k+1})}{\gamma k+1\,(O^r_{k+1})} \qquad (3.22)$$

$$\hat{C}_{r\,(k+1)} = \frac{\gamma k+1\,(T^r_{k+1}\,(Y))}{\gamma k+1\,(O^r_{k+1})} \qquad (3.23)$$

$$\sigma_{\,k+1)} = \frac{\gamma k+1\left(T^r_{k+1}(Y^2)\right) - 2_{cr}\,\gamma k+1\left(T^r_{k+1}(Y)\right) + C_r^2\,\gamma k+1(O^r_{k+1})}{\gamma k+1\,(O^r_{k+1})} \qquad (3.24)$$

iii. Write

$$\hat{A}_{(k+1)} = [\,\hat{a}_{sr\,(k+1)}] \qquad (3.25)$$

c. Calculate $\hat{\pi}\,(k+1)$ from $\hat{A}\,(k+1)\,\hat{\pi}\,(k+1) = \hat{\pi}\,(k+1)$

Repeat step a till c for the next $\ell$

Step 4.

Calculate the value of $\hat{Y}_{(k+1)} = \sum_{i=1}^{N} \hat{\pi}_i\,(k)\,\hat{C}_i\,(k)$ ............................ (3.26)

Step 5.

For $k$=1 till $T$, print $\hat{Y}_k$

(Setiawaty, 2005)

### 3.2.2 Notations

$N$ = number of states in the model

$O$ = number of observation symbols

$aij$ = state transition probabilities

$B$ = observation emission probability distribution that characterizes each state

$\pi$ = initial state distribution

$Y_k$ = observation process

$X_k$ = causes of events (state)

$C, E(x)$ = expectation value

$E[X_k \mid Y_k]$ = conditional expectation

$\sigma$ = variance

$\emptyset$ = empty set

$\mathbb{R}^N$ = real number set

$\Gamma$ = gamma function

## 3.3 Data Requirement

Data Requirement is monthly gold price per gram in rupiahs (Rp) for last three years and the causative factor.

## 3.4 Data Collecting Method

The method of collecting data is secondary data. The Gold Minted Bars data was obtained from the Logam Mulia by PT Aneka Tambang,Tbk every month from January 2008 until December 2010.

## 3.5 Data analysis

Data analysis is focused in finding the most likely sequence based on the previous model to predict the gold price. Recursive estimation is used to estimate the new parameter.

### 3.6 Tools Analysis

The model applied in data analysis is processed using Microsoft Excel-® and Matlab® to estimate the parameters based on Hidden Markov Model.

### 3.7 Frame of Research

The research steps are required to be organized properly in order to simplify the composing of research report. Figure 3.2 is the research framework.

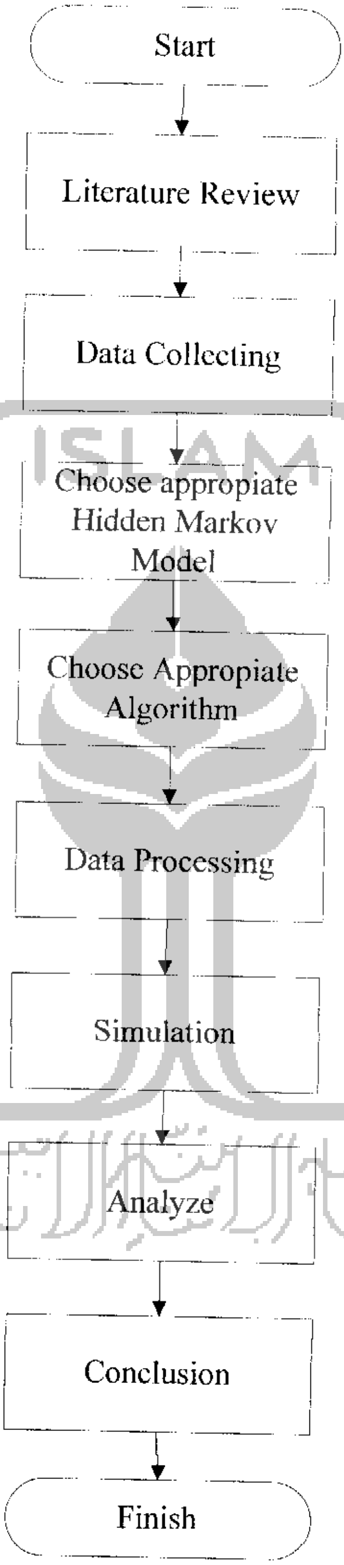


Figure 3.2 Research Framework