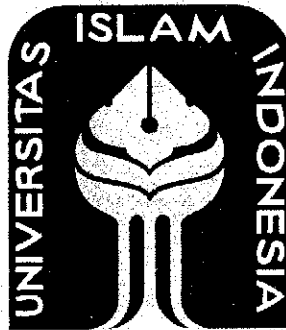


PREDICTION OF GOLD PRICE USING HIDDEN MARKOV MODEL

Thesis

Submitted to International Program

Faculty of Industrial Technology in Partial Fulfillment of
the Requirements for the Degree of Sarjana Teknik Industri
Universitas Islam Indonesia



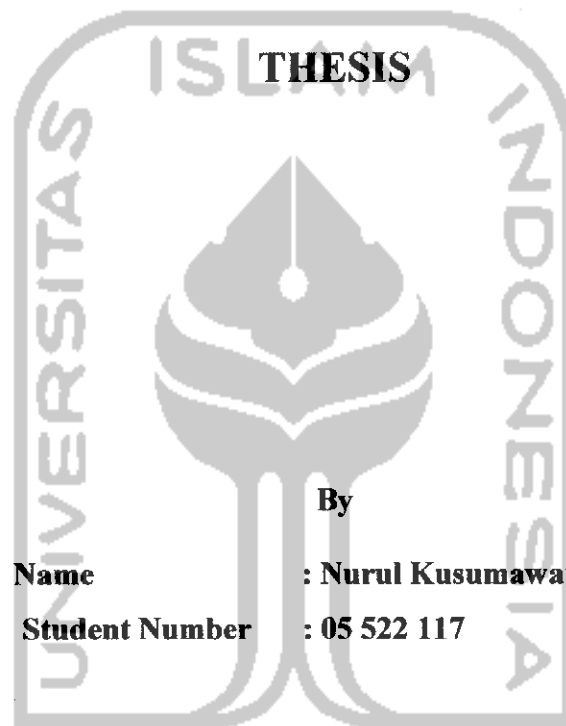
By :Nurul Kusumawardhani

Student No. : 05 522 117

**INTERNATIONAL PROGRAM
DEPARTMENT OF INDUSTRIAL ENGINEERING
FACULTY OF INDUSTRIAL TECHNOLOGY
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2011**

THESIS APPROVAL OF SUPERVISOR

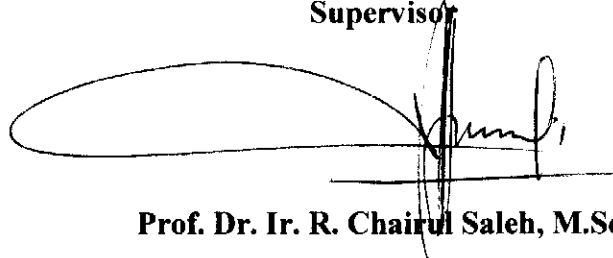
PREDICTION OF GOLD PRICE MOVEMENT USING HIDDEN MARKOV MODEL



By
Name : Nurul Kusumawardhani
Student Number : 05 522 117

Yogyakarta, November 2011

Supervisor



Prof. Dr. Ir. R. Chairul Saleh, M.Sc.

THESIS APPROVAL OF EXAMINATION COMMITTEE

PREDICTION OF GOLD PRICE USING HIDDEN MARKOV MODEL

By

Name : Nurul Kusumawardhani

Student Number : 05 522 117

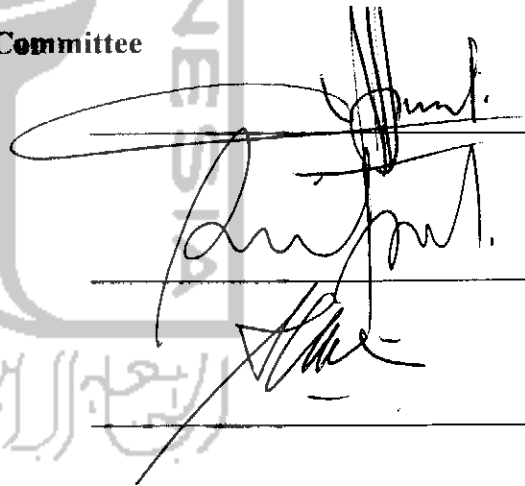
Was defended in front of Examination Committee in Partial Fulfillment of
the Requirements for the degree of Sarjana Teknik Industri
Fakultas Teknologi Industri Universitas Islam Indonesia
Yogyakarta, November 7th 2011

Examination Committee

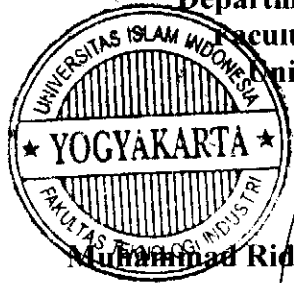
Prof. Dr. Ir. R. Chaerul Saleh, Msc
Examination Committee Chair

M Ridwan A Purnomo, ST., M.Sc., Ph.D
Member I

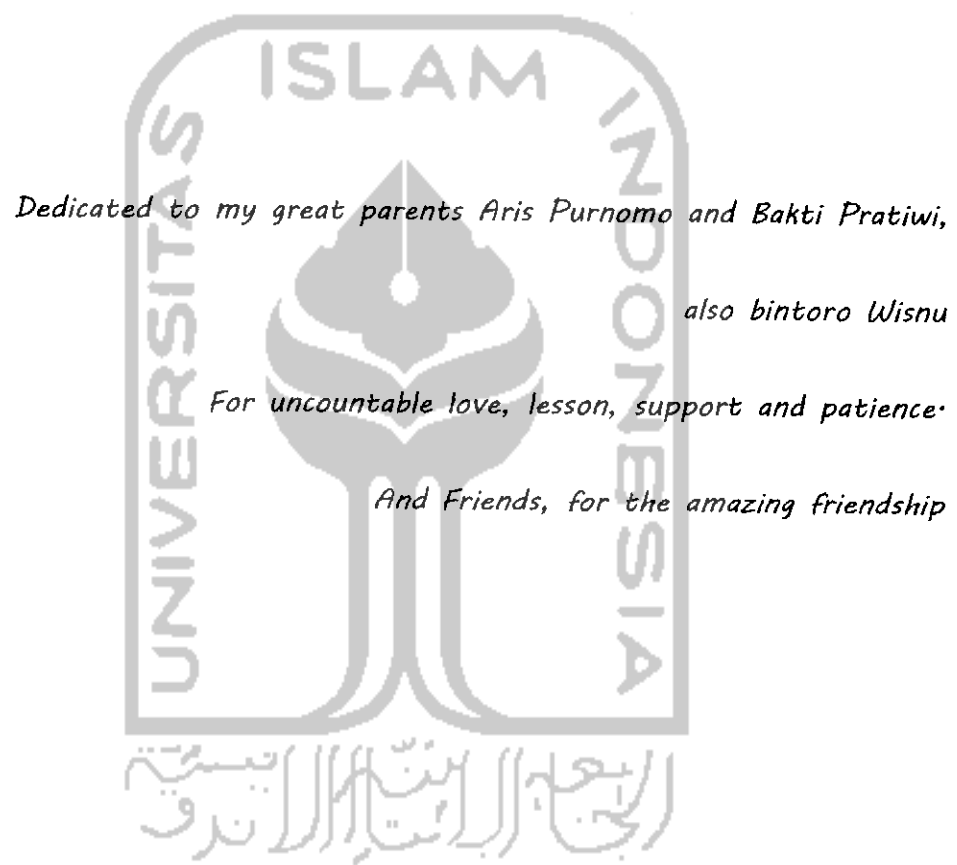
Agus Mansur ST, M Eng. SC
Member II



Accepted by,
Director of International Program
Department of Industrial Engineering
Faculty of Industrial Technology
Universitas Islam Indonesia



Muhammad Ridwan Andi Purnomo, ST., M.Sc., Ph.D



*Dedicated to my great parents Aris Purnomo and Bakti Pratiwi,
also bintoro Wisnu
For uncountable love, lesson, support and patience.
And Friends, for the amazing friendship*

ACKNOWLEDGEMENT

Assalamualaikum, Wr. Wb.

Alhamdulillah robbil 'alamin, on behave of Allah SWT this thesis report in PT Logam Mulia finished well. This report is arranged as a partial fulfillment of the requirement to have Strata One Degree at Industrial Engineering, Faculty of Industrial Technology, Universitas Islam Indonesia.

Authors know that it is not easy to make a perfect report. But because there are supports from many people, this field works eventually can be done based on the specified time. Therefore with humbly, in this time the author gives a great appreciation and thanks to:

1. Aris Purnomo and Bakti Pratiwi for the everlasting support and motivation
2. Prof. Dr. Ir. R. Chairul Saleh, M.Sc as my supervisor that always gives all directions and motivation to make this report better
3. M Ridwan Andi Purnomo, ST., Msc., Ph.D as the Director of International Program Industrial Engineering Department and IP staff for administration helps
4. Mrs Chairul and family for the kindness
5. Rizka Dian for the help that really count
6. All IP 05 for all those moments, esp Ditania, Toni, Ricky, Irfan, Alief, Adhib, Ibnu, Fadly, etc
7. Everyone who involved in finishing this report, which cannot be mentioned one by one

Author realize that in this report there are still many mistakes, and me as author still need critic and suffestion for this perfection of the report. Hopefully Allah SWT blesses this effort. Amin.

Wassalamu'alaikum Wr, Wb

Yogyakarta, November 2011

PREDICTION OF GOLD PRICE USING HIDDEN MARKOV MODEL

By:

Nurul Kusumawardhani

Submitted to International Program, Department of Industrial Engineering,
Faculty of Industrial Technology,
in partial fulfillment of the Requirements
for the degree of Sarjana Teknik Industri
at Universitas Islam Indonesia

ABSTRACT

Hidden Markov Models (HMM) is a stochastic model where the system being modeled is assumed as a Markov process with unknown parameters. The challenge is to determine the hidden parameters from observed parameters. It provides a probabilistic framework for modeling a time series forecasting. If the causative factors are not observed directly and has the properties of Markov Chain, so, a pair of observation and its causative factors is Hidden Markov Model. The proposed model was applied on minted gold price change in 2008 until 2010. Gold price was influenced by several factors, such as international recession, crude price, government policy, etc. The algorithm used to estimate the parameters, thus the estimated parameters were used to calculate the expectation value of gold to found the most likely sequence. The proposed model then would be coded using Matlab. The result of this study showed that the continuous hidden Markov model could be applied on gold with 2.25%, 2.22%, 2.96% and 2.8% of total percentage error.

Keywords: Hidden Markov Model, Time Series, Gold, Estimated Parameter

TABLE OF CONTENT

COVER	
THESIS SUPERVISOR APPROVAL	ii
THESIS EXAMINER APPROVAL	iii
DEDICATION PAGE	iv
MOTTO	v
ACKNOWLEDGEMENT	vi
ABSTRACT	vii
TABLE OF CONTENT	viii
LIST OF TABLE	x
LIST OF FIGURE	xi
CHAPTER I INTRODUCTION	
1.1 Background	1
1.2 Problem Formulation	2
1.3 Scope of Research	2
1.4 Objective of Research	3
1.5 Significance of Research	3
1.6 Systematic Writing of Thesis	3
CHAPTER II LITERATURE REVIEW	
2.1 Previous research	5
2.2 Probability Theory	
2.2.1 Probability and Random Variable	8
2.2.2 Conditional Expectation	10
2.2.3 Expected Value and Variance	11
2.3 Time Series Analysis	11
2.4 Stochastic Process	12
2.5 Markov Chain	14
2.6 Hidden Markov Model	
2.5.1 Characteristic of HMM	18
2.5.2 Main issues using HMM	18
2.7 Percentage Error	23
CHAPTER III RESEARCH METHODOLOGY	
3.1 Research Object	24
3.2 Hidden Markov Model	
3.2.1 HMM Model	24
3.2.2 Notation	30
3.3 Data Requirement	31
3.4 Data Collecting Method	31

3.5	Data Analysis	31
3.6	Tools Analysis	32
3.7	Frame of Research	32
CHAPTER IV DATA COLLECTING AND PROCESSING		
4.1	Data Collecting	34
4.2	Data Processing	40
CHAPTER V DISCUSSION		
CHAPTER VI CONCLUSION AND SUGGESTION		
6.1	Conclusion	53
6.2	Suggestion	54
REFERENCES		
APPENDICES		



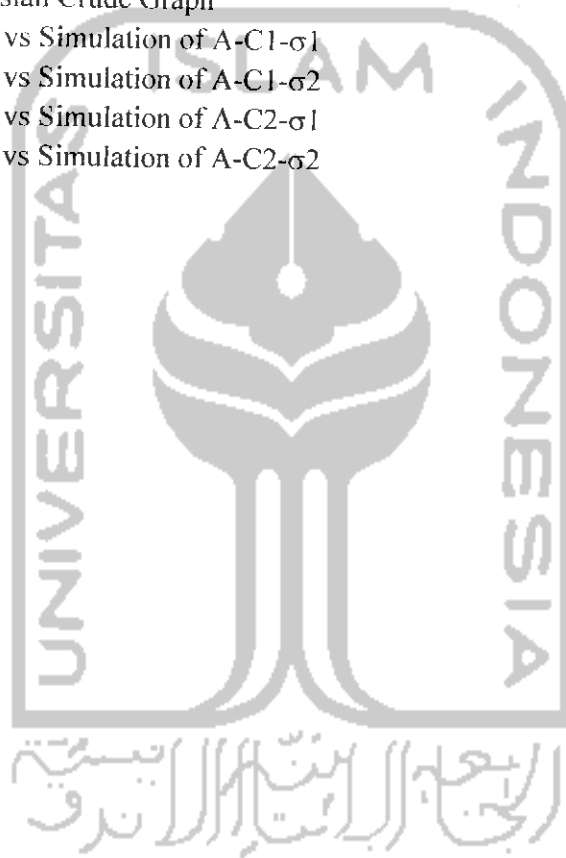
LIST OF TABLE

Table 2.1	Probability Sample Space	11
Table 4.1	Dimensions of Casting Bar	40
Table 4.2	Dimensions of Minted Bar	40
Table 4.3	Gold and Indonesian Crude Price in year 2008-2010	42
Table 4.4	State for Each Month	45
Table 4.5	State Transition	46
Table 4.6	Simulation Result of A-C1- σ_1	48
Table 4.7	Simulation Result of A-C1- σ_2	49
Table 4.8	Simulation Result of A-C2- σ_1	51
Table 4.9	Simulation Result of A-C2- σ_2	53



LIST OF FIGURE

Figure 2.1	Example of State Transition	19
Figure 2.2	Re-estimating Transition Probabilities	24
Figure 3.1	Hidden markov Model	29
Figure 3.2	Research Flowchart	37
Figure 4.1	Actual gold Graph	43
Figure 4.2	Indonesian Crude Graph	44
Figure 4.3	Actual vs Simulation of A-C1- σ_1	49
Figure 4.4	Actual vs Simulation of A-C1- σ_2	51
Figure 4.5	Actual vs Simulation of A-C2- σ_1	53
Figure 4.6	Actual vs Simulation of A-C2- σ_2	54



CHAPTER I

INTRODUCTION

1.1 Background

Probability theory can be understood as a mathematical model for the intuitive notion of uncertainty. Without probability theory all the stochastic models in Physics, Biology, and Economics would either not have been developed. These processes are the basic of classical probability theory and much of statistics. The possible outcomes for each experiment are the same and occur with the same probability when a sequence of chance experiments forms an independent trials process.

The idea of the Hidden Markov model appears to have first come out in the late 1960's at the Institute for Defense Analyses (IDA) in Princeton, New Jersey. The HMM model attempts to address the characteristics of a probabilistic sequence of observations that may not be a fixed function but instead changes according to a Markov chain. Stochastic process on the gold price, have several factors that cannot be ascertained and formed a markov chain, hence used Hidden Markov Model method to overcome the case.

Gold is one of the precious metal that has the characteristic of zero inflation. Because of its flexibility people tent to invest their money in gold recently. Naturally, gold price is formed a stochastic situation. However there are odds that it will decrease over time. Before investing their money into gold people need to know the market trend for gold itself. Therefore, they need specific method to read the trend.

According to Wai Ki Cheng et al., (2004) Hidden Markov Models have been widely used in science, engineering and many other areas and have been successfully

applied in engineering problems such as speech recognition, DNA sequences, electrical signal prediction and image processing, etc.

Previous researches for Hidden Markov Model have been done by Lukman Zaman, (2003) with On-Line Chinese Character Recognition Using Hidden Markov Models, Detection of machine failure: Hidden Markov Model approach by Allen H. Tai (2008), Application of Hidden Markov Model to Automatic Speech Endpoint Detection by Wilpon and Rabiner (1987) in which Hidden Markov Model is useful for speech reorganization. Hidden Markov Model has also been used by Yuni, (2009) in Penerapan Hidden Markov Model pada Peramalan Harga Premium.

However, the Hidden Markov Model is never applied for predicting gold price before. The research was conducted to investigate the Hidden Markov Model to be applied in time series forecasting to know the gold trend.

1.2 Problem Formulation

Based on the background, the problem formulation to be solved is how was the market trend of gold based on Hidden markov Model method?

1.3 Scope of Research

There are some boundaries in this project. There are:

- a. The research object is minted Gold price per grams in Rupiah.
- b. The domain on the research is limited for January 2008 until December 2010.
- c. The causative factor used in this research are new academic year and crude oil.

- d. Tool used in research are Matlab® and Microsoft Excel.

1.4 Objective of Research

The objective of this research is to predict the most likely hidden state sequence of gold price.

1.5 Significance of Research

After the research was being done, it will give the significance benefit:

- a. Enrich the knowledge on the application of Hidden Markov Model in prediction field.
- b. Enrich the knowledge on the application of MATLAB as mathematical programming software.
- c. As a reference for further research especially in stochastic problem.

1.6 Systematic Writing of Thesis

Systematic writing of this thesis were stated as follows.

CHAPTER I INTRODUCTION

This chapter described about the background of the study that will be the fundamental theory, problem formulation, scope of research as the boundaries, objective of research as the aim, significance of research as the benefit and systematic writing.

CHAPTER II LITERATURE REVIEW

This chapter describes the explanation of the previous researches that have been conducted at earlier time about the

thesis. It also provides the literature study that the writer is using in the process of composing the thesis. The requirement process are being used to modelled this proposed method.

CHAPTER III RESEARCH METHODOLOGY

This chapter provides explanations that consist of the object of the research, model development, framework of research and necessary data that required in completing the research which is secondary data.

CHAPTER IV DATA COLLECTING AND PROCESSING

This chapter presents information of data that have been collected during the research. It also contains the problem solving using the proposed model or tools that are implemented in the data processing.

CHAPTER V DISCUSSION

This chapter provides a discussion about the data that has been collected during the research and also the result of processed data. Furthermore, it also discuss about the result to see the ability of proposed model to solve the problems.

CHAPTER VI CONCLUSION AND RECOMMENDATION

This chapter presents the conclusions of the research that conducted and recommendation for the further research.

REFERENCE

APPENDICES

CHAPTER II

LITERATURE REVIEW

This chapter describe about the fundamental theory used to conduct the research model.

2.1 Previous research

Hassan and Nath (2005) presented Hidden Markov Models (HMM) approach for forecasting stock price at the interrelated markets. Any fluctuation in market influences personal and corporate financial lives, and the economic health of a country. Hassan et al., (2005) consider 4 input features for a stock, which is the opening price, closing price, highest price, and the lowest price. The next day's closing price is taken as the target price associated with the four input features. The idea behind new approach in using HMM is that the using of training dataset for estimating the parameter set (A, B, π) . Using the trained HMM, likelihood value for current day's dataset is calculated. For instance the likelihood value for the day is \mathcal{L} then from the past dataset using the HMM locate those instances that would produce the same \mathcal{L} or nearest to the \mathcal{L} likelihood value. Assuming that the next day's stock price should follow about the same past data pattern, from the located past day(s) simply calculate the difference of that day's closing price and next to that day's closing price. Thus the next day's stock closing price forecast is established by adding the above difference to the current day's closing price. The results show potential of using HMM for time series prediction.

Amri, (2008) discussing about Hidden Markov Model concerning about speech signal recognition using HMM-Neural Network (NN) method. The method is used to determine the sequence of speech signal data based on the initial and feature extraction from a batch of different word, NN is used to determine the success of speech signal recognition processes. In this research, a word is spelled by single utterance with different word and different number of words and being done with the set of 5-50 words in Bahasa. The condition applied is 50 number of words, NN structure used is amount to 5 layer hidden, 20 node, 10 node, 5 node, 10 node, 20 node for each.

Testing data using AT & T Database by the number of 40 individuals with three training images per individual and 7 images per individual testing. While the Yale Face Database with a number of 15 individuals with 4 images per individual and 7 training images per individual testing. Those models have successfully built applications embedded HMM-based face recognition with identification accuracy of 94.64% generalization (AT & T Database) and 77.14% (Yale Face Database). By the results it is known that AT & T database that does not require cutting face area is much better than the Yale Face Database. This is because the cutting area of the face in an image depends on the accuracy of face detection, so changes will take effect in the detection area of training and testing results.

Irfani et al., (2006) discussed about speech recognition. The development of speech recognition technology is one form of technological developments in the 20th century that utilize voice as input. The voice is an alternative method for humans to interact with computers. The computer will recognize the voice commands and perform the stretcher as a reaction to the command. Modern speech recognition systems are generally based on

Hidden Markov Models (HMMs). By HMMs sound signals can be characterized as a random process parameters, and parameters of the stochastic process which can be determined precisely. That statistical model then processed using the Viterbi algorithm. Viterbi algorithm is a dynamic programming algorithm to find possible hidden state sequence (commonly called the Viterbi path) which produced the series of observations of events, especially within the scope of the HMM. By viterbi algorithm processing statuses in the voice recognition system can be optimized.

On the other hand, the expansion of Hidden markov model was also discussed by Imam (2007). In this research, an application for face recognition is built based on embedded Hidden Markov Models (eHMM). eHMM are able to modeling image as 2 dimensional data better than ordinary HMM. eHMM will segmented face area on digital image into 5 super states (forehead, eyes, nose, lips, and chin) and into couple of embedded states inside those super states. Identifications are being done by comparing test image's observation likelihood with eHMM face model. Viterbi Algorithm is used to evaluate the best likelihood from the comparison of test image's observation with all people's eHMM. The accuracy of identification is known by introduction to each test image belongs to every individual.

The other study has been done by Fathoni, (2008) in Continuous hidden markov model review and its application on harvest unhusked. The harvest unhusked price is fluctuated according to some factors such as thai-rice import policy, fertilizer, harvest failure, natural disaster, political situation etc. Those factors assumed as the state of unobservable markov chain. The input data was the average at the producers in region I from January 2000 through March 2007. It is assumed that the price of unhusked raised

by the random variables Y_k that spread in a certain distribution at interval changes (Ω, \mathcal{F}, P) . To facilitate the search for the parameter estimators created a functional programming-based computing using Mathematica 6.0. The estimators obtained are used to calculate the expected value of the price of unhusked rice. From the results obtained, the continuous hidden Markov model is good enough explain the behavior of the price harvest. The more the cause value of the alleged incident, the better.

Research in the field of prediction using Hidden markov model is a new breakthrough, because most of them are used for bio technology or recognition. Previous literature suggests that there has been no research on the gold market scientifically. Community observe the trends based on intuition and a simple calculation. Meanwhile they need to know the market trend for gold itself. This research was intended to study time series trend using Hidden Markov Model to give a scientific description by finding the most likely sequence.

2.2 Probability theory

The fields of statistics are related to the ways of data collection, processing, presentation, analytical and conclusion has been made based on data and analysis. Resulting conclusion is expected to be the description of population and their characteristics. Experiment is a method of data collecting. Set of all possible outcomes from a randomized trial called the sample space and its denoted by Ω . An event A is a subset of the sample space Ω .

2.2.1 Probabilty And Random Variabel

If an experiment obtain a continued sample, then the random variable that connected to those sample is called continued random variable. The probability spread of

random variable is called probability density function (Lungan, 2006). A random variable x is a function of $x: \Omega \rightarrow \mathbb{R}$ which $\{ \omega \in \Omega : X(\omega) \leq X \} \in \text{field}$, for each $X \in \mathbb{R}$. It is not a variable but rather a function that maps events to numbers. A Random Variable is a function, which assigns unique numerical values to all possible outcomes of a random experiment under fixed conditions (Ali, 2000).

This example is extracted from Ali, (2000). Suppose that a coin is tossed three times and the sequence of heads and tails is noted. The sample space for this experiment evaluates to: $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$. Now let the random variable X be the number of heads in three coin tosses. X assigns each outcome in S a number from the set $S_x = \{0, 1, 2, 3\}$. The table below lists the eight outcomes of S and the corresponding values of X .

Table 2.1 Probability sample space

Outcome	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
X	3	2	2	1	2	1	1	0

X is then a random variable taking on values in the set $S_x = \{0, 1, 2, 3\}$. Mathematically, a random variable is defined as a measurable function from a probability space to some measurable space. This measurable space is the space of possible values of the variable, and it is usually taken to be the real numbers. The condition for a function to be a random variable is that the random variable cannot be multivalued. There are 2 types of random variables:

- a. A Continuous Random Variable is one that takes an infinite number of possible values. Example: Duration of a call in a telephone exchange.

b. A Discrete Random Variable is one that takes a finite distinct value.

Example: A number of students who fail a test.

a. Continuous random variable

If the random variable X is continuous with probability density function $f(x)$,

$$\text{Var}(X) = \int (x-\mu)^2 f(x) dx, \quad (2.1)$$

where μ is the expected value, i.e.

$$\mu = \int x f(x) dx, \quad (2.2)$$

b. Discrete random variable

If the random variable X is discrete with probability mass function $x_i \rightarrow p_1, \dots, x_n \rightarrow p_n$, then

$$\text{Var}(X) = \sum_{i=1}^n p_i \cdot (x_i - \mu)^2 \quad (2.3)$$

where μ is the expected value, i.e.

$$\mu = \sum_{i=1}^n p_i \cdot x_i \quad (2.4)$$

2.2.2 Conditional Expectation

In probability theory, a conditional expectation, also known as conditional expected value or conditional mean, is the expected value of a real random variable respect to a conditional probability distribution.

2.2.3 Expected Value and Variance

The expected value, or mean, of a random variable is the weighted average of all possible values. The weights used in computing this average correspond to the probabilities in case of a discrete random variable, or densities in case of a continuous random variable. In probability theory and statistics, the variance is used as a measure of how far a set of numbers are spread out from each other. It is describing how far the numbers lie from the mean (expected value).

The variance is a parameter describing the actual probability distribution of an observed population. In the latter case a sample of data from a distribution can be used to construct the variance defined below.

If a random variable X has the expected value (mean) $\mu = E[X]$, then the variance of X is given by:

$$\text{var}(X) = E[(X - \mu)^2] \dots\dots\dots (2.5)$$

2.3 Time Series Analysis

A time series is a sequence of observations of a random variable. Hence, it is a stochastic process. Examples include the monthly demand for a product or the annual freshman enrollment in a department of a university. Forecasting time series data is important component of operations research because these data often provide the foundation for decision models. An inventory model requires estimates of future demands and a course scheduling and staffing model for a university requires estimates of future student inflow. Time series analysis provides tools for selecting a model that can be used to forecast of future events. Modeling the time series is a statistical problem. Forecasts are used in

computational procedures to estimate the parameters of a model being used to allocated limited resources or to describe random processes. There are two main goals of time series analysis:

- a. Identifying the nature of the phenomenon represented by the sequence of observations.
- b. Predicting future values of the time series variable.

Both of these goals require that the pattern of observed data is identified and described. Once the pattern is established, we can interpret and integrate it with other data.

2.4 Stochastic Process

A probability space associated with a random experiment is a triple (Ω, \mathcal{F}, P) where:

- (i) Ω is the set of all possible outcomes of the random experiment, and it is called the sample space.
- (ii) \mathcal{F} is a family of subsets of Ω which has the structure of a σ field:
 - a) $\emptyset \in \mathcal{F}$
 - b) If $A \in \mathcal{F}$, then its complement A^c also belongs to \mathcal{F}
 - c) $A_1, A_2, \dots, \in \mathcal{F} \rightarrow \bigcup_{t=1}^{\infty} A_t \in \mathcal{F}$
- (iii) P is a function which associates a number $P(A)$ to each set $A \in \mathcal{F}$ with the following properties:

- a) $0 \leq P(A) \leq 1$,
- b) $P(\Omega) = 1$

c) For any sequence A_1, A_2, \dots of disjoint sets in \mathcal{F} (that is, $A_i \cap A_j = \emptyset$ if $i \neq j$)

$$d) P\left(\bigcup_{t=1}^{\infty} A_t\right) = \sum_{t=1}^{\infty} P(A_t)$$

The elements of the σ -field \mathcal{F} are called events and the mapping P is called a probability measure. In this way we have the following interpretation of this model:

$$P(F) = \text{probability that the event } F \text{ occurs}$$

The set \emptyset is called the empty event and it has probability zero. Indeed, the additivity property (iii,c) implies:

$$P(\emptyset) + P(\emptyset) + \dots = P(\emptyset)$$

The set Ω is also called the certain set and by property (iii,b) it has probability one. Usually, there will be other events $A \in \mathcal{F}$ such that $P(A) = 1$. If a statement holds for all ω in a set A with $P(A) = 1$, then the statement is true, or that the statement holds for almost all $\omega \in \Omega$. The axioms a), b) and c) lead to the following basic rules of the probability calculus:

$$P(A \cup B) = P(A) + P(B) \text{ if } A \cap B = \emptyset$$

$$P(A^c) = 1 - P(A)$$

$$A \subset B \rightarrow P(A) \leq P(B)$$

Example : Consider the experiment of flipping a coin once.

$$\Omega = \{H, T\} \text{ (the possible outcomes are "Heads" and "Tails")}$$

$$\mathcal{F} = P(\Omega) \text{ (}\mathcal{F} \text{ contains all subsets of } \Omega\text{)}$$

$$P(\{H\}) = P(\{T\}) = \frac{1}{2}$$

2.5 Markov Chain

Andrey Markov produced the first results (1906) for these processes theoretically. Markov chains are related to Brownian motion and the ergodic hypothesis, but Markov appears to have pursued this out of a mathematical motivation, namely the extension of the law of large numbers to dependent events. In 1913, he applied his findings for the first time to the first 20,000 letters of Pushkin's Eugene Onegin.

A Markov chain is a discrete random process with the property that the next state depends only on the current state. It is named for Andrey Markov, and is a mathematical tool for statistical modeling in modern applied mathematics, particularly information sciences. A useful heuristic is that of a frog jumping among several lily-pads, where the frog's memory is short enough that it doesn't remember what lily-pad it was last on, and so its next jump can only be influenced by where it is now.

Formally, a Markov chain is a discrete random process with the Markov property that goes on forever. A discrete random process means a system which is in a certain state at each step, with the state changing randomly between steps. The steps are often thought of as time (such as in the frog and lily-pad example), but they can equally well refer to physical distance or any other discrete measurement. The Markov property states that the conditional probability distribution for the system at the next step depends only on the current state of the system, and not additionally on the state of the system at previous steps:

$$P(X_{n+1}|X_1, X_2, \dots, X_n) = P(X_{n+1}|X_n) \dots\dots\dots (2.6)$$

Since the system changes randomly, it is generally impossible to predict the exact state in the future. However, the statistical properties of the system's future can be predicted. In many applications it is these statistical properties that are important. The changes of state are called transitions, and the probabilities associated with various state-changes are called transition probabilities. The set of all states and transition probabilities completely characterizes a Markov chain. By convention, assume all possible states and transitions have been included in the definition of the processes, so there is always a next-state and the process goes on.

A famous Markov chain is the drunkard's walk, a random walk on the number line where the position may change by +1 or -1 with equal probability. From any position there are two possible transitions, to the next or previous integer. The transition probabilities depend only on the current position, not on the way the position was reached. For example, the transition probabilities from 5 to 4 and 5 to 6 are both 0.5, and all other transition probabilities from 5 are 0. These probabilities are independent of whether the system was previously in 4 or 6.

However, the theory is usually applied only when the probability distribution of the next step depends non-trivially on the current state. A Markov chain is a sequence of random variables X_1, X_2, X_3, \dots with the Markov property, given the present state, the future and past states are independent. Formally,

$$\Pr(X_{n+1}=\chi|X_1=\chi_1, X_2=\chi_2, \dots, X_n=\chi_n)=\Pr(X_{n+1}=\chi|X_n=\chi_n) \dots \dots \dots (2.7)$$

The possible values of X_i form a countable set S called the state space of the chain. Markov chains are often described by a directed graph, where the edges are labeled by the

probabilities of going from one state to the other states. A simple example is shown in the figure below, using a directed graph to picture the state transitions. The states represent whether the economy is in a bull market, a bear market, or a recession, during a given week. According to the figure, a bull week is followed by another bull week 90% of the time, a bear market 7.5% of the time, and a recession the other 2.5%. From this figure it is possible to calculate, for example, the long-term fraction of time during which the economy is in a recession, or on average how long it will take to go from a recession to a bull market.

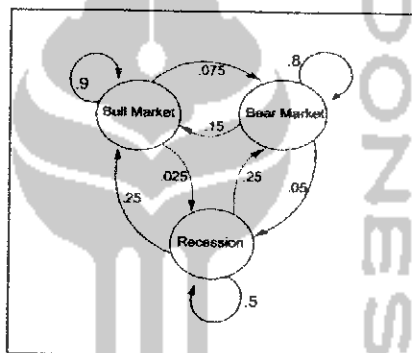


Figure 2.1 Example of state transition

The probability of going from state i to state j in n time steps is

$$P_{ij} = \Pr (X_n = j | X_0 = i) \dots\dots\dots (2.8)$$

and the single-step transition is

$$P_{ij} = \Pr (X_1 = j | X_0 = i) \dots\dots\dots (2.9)$$

For a time-homogeneous Markov chain:

$$P_{ij} = \Pr (X_{n+k} = j | X_k = i) \dots\dots\dots (2.10)$$

And

$$P_{ij} = \Pr(X_{k+1} = j | X_k = i) \dots\dots\dots (2.11)$$

so, the n -step transition satisfies the Chapman–Kolmogorov equation, that for any k such that $0 < k < n$,

$$P_{ij}^{(n)} = \sum_{r \in S} p_{ir}^{(k)} p_{rj}^{(n-k)} \dots\dots\dots (2.12)$$

The marginal distribution $\Pr(X_n = x)$ is the distribution over states at time n . The initial distribution is $\Pr(X_0 = x)$.

2.6 Hidden Markov Model

A Hidden Markov Model (HMM) is a finite state machine which has some fixed number of states. Hidden Markov models were introduced in the beginning of the 1970's as a tool in speech recognition. This model based on statistical methods has become increasingly popular in the last several years due to its strong mathematical structure and theoretical basis for use in a wide range of applications. If the parameters of the chain are known, quantitative predictions can be made. In other cases, they are used to model a more abstract process, and the theoretical underpinning of an algorithm. Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state an outcome or observation can be generated, according to the associated probability distribution. It is only the outcome, not the state visible to an external observer and therefore states are hidden to the outside, hence the name Hidden Markov Model

2.6.1 Characteristic of Hidden Markov Model

Hidden Markov Model is characterized by the following

- a. number of states in the model (N / Q)

$$Q = \{q_1; q_2; \dots; q_T\} \text{ - set of states}$$

- b. number of observation symbols (M / O)

$$O = \{o_1; o_2; \dots; o_T\} \text{ - set of symbols}$$

- c. state transition probabilities (a_{ij})

$$a_{ij} = P(q_{t+1} = j | q_t = i)$$

- d. observation emission probability distribution that characterizes each state (b_j)

$$b_j(k) = P(o_t = k | q_t = j) \quad i \leq k \leq M$$

- e. initial state distribution (π)

As mentioned above the HMM is characterized by N , M , A , B and π . The a_{ij} , $b_j(O_t)$, and π_i have the properties:

$$\sum_j a_{ij} = 1, \quad \sum_t b_i(O_t) = 1, \quad \sum_i \pi_i = 1 \quad \text{and}$$

$$a_{ij}, b_j(O_t), \pi_i \geq 0 \text{ for all } i, j, t. \quad \dots \dots \dots (2.13)$$

2.6.2 Main issues using HMM

Most applications of HMMs are finally reduced to solving three main problems. These are:

a. Evaluation problem

Given the HMM $\lambda = (A, B, \pi)$ and the observation sequence $O = o_1 o_2 \dots o_K$, calculate the probability that model M has generated sequence O . Trying to find probability of observations $O = o_1 o_2 \dots o_K$ by means of considering all hidden state sequences.

b. Decoding problem

Given the HMM $\lambda = (A, B, \pi)$ and the observation sequence $O = o_1 o_2 \dots o_K$, calculate the most likely sequence of hidden states s_i that produced observation sequence O .

c. Learning problem

Given some training observation sequences $O = o_1 o_2 \dots o_K$ and general structure of HMM (numbers of hidden and visible states), determine HMM parameters $\lambda = (A, B, \pi)$ that best fit training data.

$$a_{ij} = P(s_i | s_j) = \frac{\text{number of transition from state } S_j \text{ to state } S_i}{\text{number of transitions out of state } S_j}$$

$$b_j(v_m) = P(v_m | s_j) = \frac{\text{number of time observation } V_m \text{ occurs in state } S_i}{\text{Number of times in state } S_i}$$

$$\pi_i = P(s_i) = \text{Expected frequency in state } s_i \text{ at time } k=1$$

In a Hidden Markov Model, three parameters need to be re-estimated, which is:

- a. Transition probabilities (a_{ij})
- b. Initial state distribution (π)

c. Emission probabilities $[b_j(o_t)]$

a. Re-estimating Transition Probabilities

What's the probability of being in state s_i at time t and going to state s_j , given the current model and parameters?

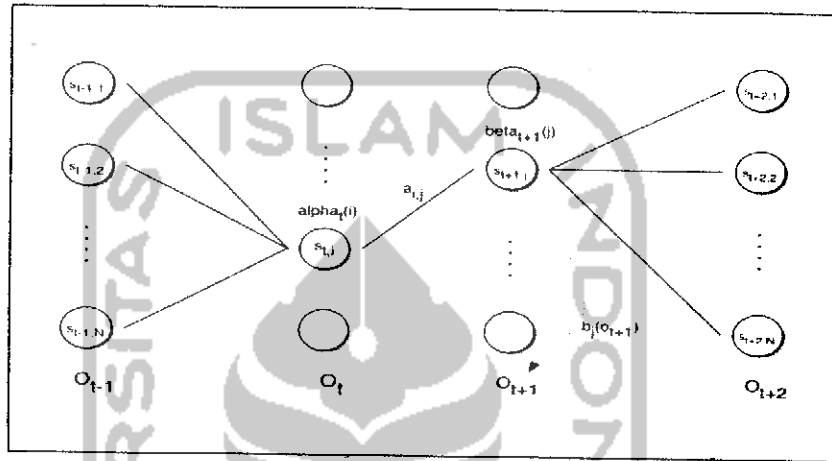


Figure 2.2 Reestimating transition probabilities

Given :

$$\xi_t(i, j) = P(q_t = s_i, q_{t+1} = s_j | O, \lambda)$$

So that,

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{i,j} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{i,j} b_j(o_{t+1}) \beta_{t+1}(j)} \quad (2.14)$$

The intuition behind the re-estimation equation for transition probabilities is:

$$\hat{a}_{i,j} = \frac{\text{expected number of transitions from state } s_i \text{ to state } s_j}{\text{expected number of transitions from state } s_i}$$

$$= \hat{a}_{i,j} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \sum_{j'=1}^N \xi_t(i,j')} \dots\dots\dots (2.15)$$

Defining

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i,j) \dots\dots\dots (2.16)$$

As the probability of being in state s_i , given the complete observation O and can be written as:

$$\hat{a}_{i,j} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \dots\dots\dots (2.17)$$

b. Re-estimation of Initial State Probabilities

Initial state distribution π_i is the probability that s_i is a start state

$\hat{\pi}_i$ = expected number of state s_i at time 1

Thus can be written:

$$\hat{\pi}_i = \gamma_1(i) \dots\dots\dots (2.18)$$

c. Re-estimation of Emission Probabilities

Emission probabilities are re-estimated as:

$$\hat{b}_i(k) = \frac{\text{expected number of times in states } s_i \text{ and observesymbol } v_k}{\text{expected number of times in states } s_i}$$

$$\hat{b}_i(k) = \frac{\sum_{t=1}^T \delta(o_t, v_k) \gamma_t(i)}{\sum_{t=1}^T \gamma_t(i)} \quad \dots \quad (2.19)$$

Where $\delta(o_t, v_k) = 1$, if $o_t = v_k$, and 0 otherwise

d. Updated Model

Coming from $\lambda = (A, B, \pi)$ we get to $\lambda' = (\hat{A}, \hat{B}, \hat{\pi})$ by the following update rules:

$$\hat{a}_{i,j} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad , \quad \dots \quad (2.20)$$

$$\hat{b}_i(k) = \frac{\sum_{t=1}^T \delta(o_t, v_k) \gamma_t(i)}{\sum_{t=1}^T \gamma_t(i)} \quad , \quad \dots \quad (2.22)$$

$$\hat{\pi}_i = \gamma_1(i) \quad , \quad \dots \quad (2.23)$$

2.6 Percentage Error

The percentage error for each number obtained by Percentage Mean Error (PME) stated below:

$$\left(\sum_{k=1}^T \frac{(Y_k - \hat{Y}_k)}{Y_k} \times 100\% \right) \dots\dots\dots (2.23)$$

Where

Y_k = actual data

\hat{Y}_k = simulation data



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:

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

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 $P(\text{'Low'}|\text{'Low'}) = 0.3$; $P(\text{'High'}|\text{'Low'}) = 0.7$;
 $P(\text{'Low'}|\text{'High'}) = 0.2$ and $P(\text{'High'}|\text{'High'}) = 0.8$
- c. Observation probabilities from $P(\text{'Rain'}|\text{'Low'}) = 0.6$; $P(\text{'Dry'}|\text{'Low'}) = 0.4$;
 $P(\text{'Rain'}|\text{'High'}) = 0.4$ and $P(\text{'Dry'}|\text{'High'}) = 0.6$
- d. Initial probabilities say $P(\text{'Low'}) = 0.4$, $P(\text{'High'}) = 0.6$

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

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

Where first term is:

$$\begin{aligned} &P(\{\text{'Dry'}, \text{'Rain'}\}, \{\text{'Low'}, \text{'Low'}\}) \\ &= P(\{\text{'Dry'}, \text{'Rain'}\} | \{\text{'Low'}, \text{'Low'}\}) P(\{\text{'Low'}, \text{'Low'}\}) \\ &= P(\text{'Low'}) P(\text{'Dry'}|\text{'Low'})P(\text{'Rain'}|\text{'Low'}) P(\text{'Low'}|\text{'Low'}) \\ &= 0.4 * 0.4 * 0.6 * 0.3 \end{aligned}$$

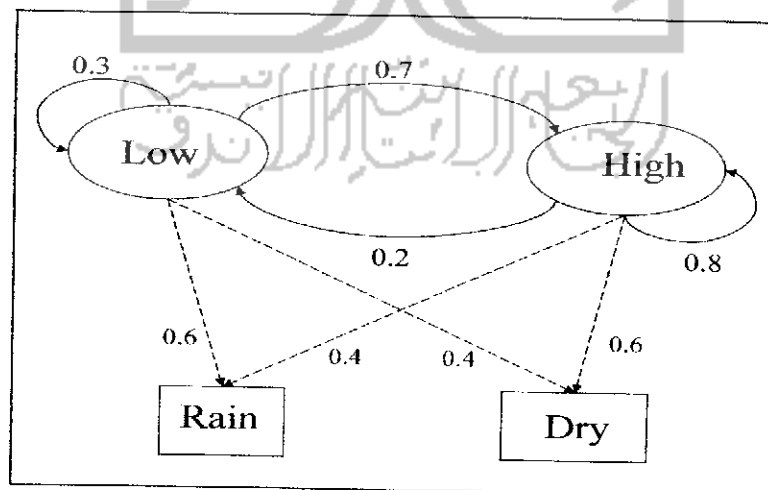


Figure 3.1 Hidden Markov Model

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 (Ω, \mathcal{F}, P) . State equations and observation equations for the second model is denoted below:

$$\begin{aligned} X_{k+1} &= AX_k + V_{k+1} \\ Y_{k+1} &= C(X_k) + \sigma_{(k+1)} \omega_{k+1} \end{aligned} \quad (3.1)$$

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 σ 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 $1 \leq i \leq N$.

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

$$P (Y_{k+1} \leq t | Y_k) = \sum_{i=1}^N \langle X_k, e_i \rangle \int_{-\infty}^{-c_i} \phi_i(x) dx \dots\dots\dots (3.2)$$

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

$$\sum_{j=1}^N \langle X_k, e_j \rangle \phi_j(t-c_j) \dots\dots\dots (3.3)$$

3. While the mutual distribution is

$$P(X_k=e_i, Y_{k+1} \leq t | Y_k) = \langle X_k, e_i \rangle \int_{-\infty}^{-c_i} \phi_i(x) dx \dots\dots\dots (3.4)$$

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

$$E [X_k | Y_{k+1}] = \frac{\sum_{i=1}^N \langle X_k, e_i \rangle \phi_i (Y_{k+1}-c_i) e_i}{\sum_{j=1}^N \langle X_k, e_j \rangle \phi_j (Y_{k+1}-c_j)} \dots\dots\dots (3.5)$$

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

$$E [X_{k+1} | Y_{k+1}] = \frac{\sum_{i=1}^N \langle X_k, e_i \rangle \phi_i (Y_{k+1}-c_i) \Lambda e_i}{\sum_{j=1}^N \langle X_k, e_j \rangle \phi_j (Y_{k+1}-c_j)} \dots\dots\dots (3.6)$$

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 $\gamma_{k+1} (X_{k+1}) = \sum_{i=1}^N \langle \gamma_k (X_k), \Gamma^i (Y_{k+1}) \rangle a_i \dots\dots\dots (3.7)$

2. Number of leap estimator

A markov chain moved out of state e_r at time k to state e_s 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

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

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

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:

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

4. Observation process estimator

The estimator for observation process will be defined as follow:

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

From the model the algorithm will be written as follow:

Step 1.

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

Step 2.

Determine the initial value

$$\pi = (\pi_i)_{2 \times 1}$$

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

$$C = (C_i)_{2 \times 1} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

$$\sigma_l = (\sigma_i)_{2 \times 1} = \begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix} \text{ with } \pi = E[X_0] \text{ that fulfill } A_\pi = \pi \text{ and } \sum_{i=1}^N \pi_i = 1$$

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)

$$\gamma_0 = (X_0) = \pi$$

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

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

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

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

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

i. Calculate recursive estimators

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

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

$$es \dots (3.13)$$

$$\gamma_{k+1}(J_{k+1}^{TS}) = \langle \gamma_{k+1, k+1}(J_{k+1}^{TS}), \mathbb{1} \rangle \dots (3.14)$$

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

$$\dots (3.15)$$

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

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

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

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

$$\text{with } \Gamma^{(i)}(Y_k) = \frac{\phi \frac{Y_k - c(i)}{\sigma(i)}}{\sigma(i)\phi(Y_k)} e^{(\cdot)} \dots (3.19)$$

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

$$\gamma_{k+1}(T_{k+1}|T_k) = \gamma_{k+1,k+1}(T_{k+1}) \dots\dots\dots (3.20)$$

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

ii. Calculate parameter estimators

$$\hat{\alpha}_{sr(k+1)} = \frac{\gamma_{k+1}(J_{k+1}^{rs})}{\gamma_{k+1}(O_{k+1}^r)} \dots\dots\dots (3.22)$$

$$\hat{C}_r(k+1) = \frac{\gamma_{k+1}(T_{k+1}^r(Y))}{\gamma_{k+1}(O_{k+1}^r)} \dots\dots\dots (3.23)$$

$$\sigma_{k+1} = \frac{\gamma_{k+1}(T_{k+1}^r(Y^2)) - 2c_r \gamma_{k+1}(T_{k+1}^r(Y)) + C_r^2 \gamma_{k+1}(O_{k+1}^r)}{\gamma_{k+1}(O_{k+1}^r)} \dots\dots\dots (3.24)$$

iii. Write

$$\hat{A}(k+1) = [\hat{\alpha}_{sr(k+1)}] \dots\dots\dots (3.25)$$

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 ℓ

Step 4.

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

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
- a_{ij} = state transition probabilities
- B = observation emission probability distribution that characterizes each state
- π = initial state distribution

Y_k = observation process

X_k = causes of events (state)

$C, E(x)$ = expectation value

$E [X_k | Y_k]$ = conditional expectation

σ = variance

\emptyset = empty set

\mathbb{R}^N = real number set

Γ = 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.



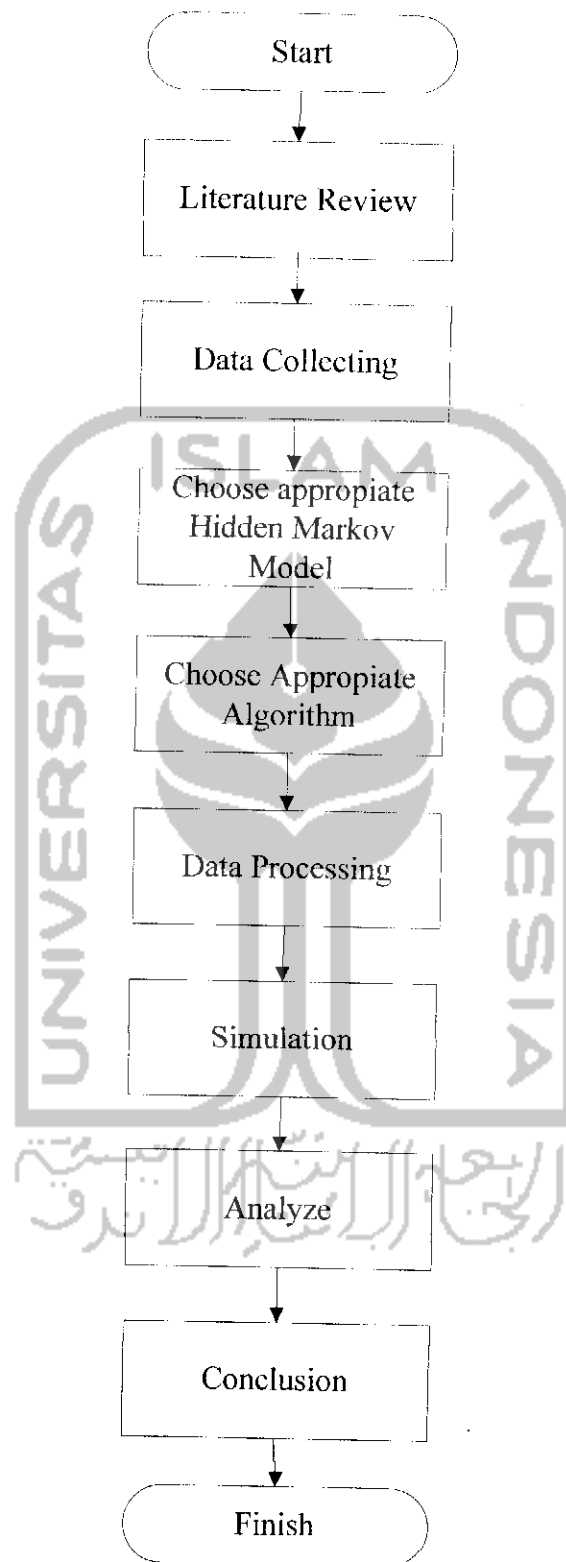


Figure 3.2 Research Framework

CHAPTER IV

DATA COLLECTING AND PROCESSING

4.1 Data Collecting

4.1.1 Logam Mulia Business Unit

Logam mulia business unit by PT Aneka Tambang (Persero) Tbk located in Jl. Pemuda - Jl. Raya Bekasi KM.18. Pulogadung, Jakarta is one of refining gold and silver in Indonesia. It has the recognition of London Bullion Market Associates (LBMA) and included in Good Delivery List of acceptable refiners of gold bars since January 1, 1999.

The company was founded by RT Braakensiek, a gold merchant around 1930 who started his business by refining gold scrap in certain locations of work. In 1937 he settled his business on Jalan Gajah Mada 84, Jakarta and began to refine gold from Bengkulu and Cikotok gold mine besides gold scrap.

In the year 1949 this private enterprise turned into a legal entity called NV Essaieur en Affinage Bedrijf v/h RT Braakensiek. In 1957 the company was taken over by Bank Industri Negara and changed its name into PT Logam Mulia (Logam Mulia Ltd). Then on January 1, 1961 under PP No.281/1961, became a state owned company named PN Logam Mulia. On December 31, 1974 in accordance with PP No.320/1974 it became one of production units PT Aneka Tambang named Unit Pengolahan dan Pemurnian Logam Mulia (Precious Metal

Refinery) and since April, 1979 the plant was then relocated to Pulogadung in East Jakarta.

Logam Mulia main business recently depends on dore bullion's supplies from several national mining companies and pongkor gold mine business unit. The maximum production capacity of Logam Mulia itself actually around 60 tons of gold and 250 tons of silver per year. But from that maximum capacity, now we only use about 60% of it and still try to increase due to the upcoming opening of new gold mines or traditional mines.

The main product of Logam Mulia is cast bars, minted bars, dinar and dirham coins and industrial products i.e needle test, silver granules, silver nitrate etc.

4.1.2 Logam Mulia Gold Flagship Product

99.99% pure gold standard is the raw material for 14 to 24 carats gold jewelry in which the levels are equivalent to the gold content. Mixing metal for the Gold Jewelry is silver and copper. Gold with a slightly reddish appearance is the dominant of copper while the yellowish appearance is dominant of silver. There is a certificate of assurance while investing in precious metal gold and can be traded anywhere.

For the consistency of long-term growth, the experts repeatedly urged the investors to diversify their investment into gold. Gold has produced many other types of investments for long-term investment category, and this is a classical protection to against inflation, holding value in the instability fluctuations of money.

Purchasing Precious Metals Gold, meaning we have invested a sum of money which the value will continue to be strengthened from time to time. When investing in the stock market plagued by uncertainty due to the inflated stock value or plummeted, gold remains strong as a solid investment tools because the value of gold does not depend on the success or failure of a firm or industry. Gold is an asset that determines its own merits. It is real, portable, can be availed anytime and anywhere.

For those varieties of advantages, Logam Mulia presents the precious gold with the weight and guaranteed content, as well as 99,99% gold purity. Gold is flexible and accepted throughout the world without need to be examined more, because it has been bundled by the LBMA Certificate (London Bullion Market Association). Gold Bullion by LM is available in gold casting, and minting. The physical appearance of casting bar will be defined below:

Table 4.1 Dimensions of Casting Bar

weight (gram)	thickness (mm)	dimension
12500 (400 troy oz)	42	238 x 78
1000	8.85	118 x 53.7
250	3.11	83.3 x 50

While the minted bars is:

Table 4.2 Dimensions of Minted Bars

weight (gram)	length (mm)	dimension
100	3.73	50 x 30
50	2.53	42.5 x 25.5
25	2.07	33.3 x 20
10	1.22	27.5 x 16.5
5	1.09	20.5 x 12.3

4	0.83	20.5 x 12.3
3	1.03	16.7 x 10
2.5	0.83	16.7 x 10
2	0.75	15.6 x 9.4
1	0.51	13 x 7.8

4.1.3 Gold Calculation for Rupiah

The troy ounce (ozt) is a unit of imperial measure. Currently this is the most common unit used to gauge the weight of precious metals. The troy ounce is part of the troy weights system derived from the Roman monetary system. The Romans used bronze bars of varying weights as currency. An *aes grave* weighed equal to 1 pound. One twelfth of an *aes grave* was called an *uncia*, or "ounce". One troy ounce defined as exactly $0.0311034768 \text{ kg} = 31.1934768 \text{ gr}$.

Since Indonesia use Rupiah for the national currency, it needs to convert the price from dollar into rupiah, which is:

$$\text{Rp / gram} = \frac{\text{loco price}}{31.103} \times \text{rate US dollar}$$

For example:

$$\text{Loco price} = 1400 \text{ USD per troy ounce}$$

$$\text{USD rate} = \text{Rp.8700,-}$$

Therefore price in rupiahs per grams will be:

$$\text{Rp / gram} = \frac{1400}{31.103} \times 8700$$

$$= \text{Rp.391.639,87 / gram}$$

4.1.4 Data Collecting

This research use secondary data obtained from Logam Mulia. Gold price in a specified period is used as data representing a time series situation. Defined:

State 1 = increasing price

State 2 = decreasing price

United States recession, the shift in crude price, new academic term in July, government policy are some example of causative factors. In this case the Indonesian crude price and new academic term are used as the factors. Actual gold price data per gram (in Rupiahs) and Indonesian Crude price year 2008, 2009 and 2010 are shown in Table 4.3.

Table 4.3 Gold and Indonesian Crude Price in year 2008-2010

Month	Actual Gold Price (Rp/gr)	ICP (US\$/barrel)
Jan-08	250000	92.09
Feb-08	298000	94.64
Mar-08	308500	103.11
Apr-08	315000	109.30
May-08	320000	124.67
Jun-08	292000	132.36
Jul-08	285500	134.96
Aug-08	295500	135.56
Sep-08	308500	99.06
Oct-08	283500	70.66
Nov-08	335000	72.32
Dec-08	350500	52.45
Jan-09	330500	61.89
Feb-09	367000	65.10
Mar-09	346500	43.95
Apr-09	330000	40.62
May-09	310000	57.86
Jun-09	319000	68.91
Jul-09	311500	64.85
Aug-09	309000	72.47

Sep-09	319500	67.07
Oct-09	334500	72.53
Nov-09	375500	77.08
Dec-09	350000	75.58
Jan-10	338500	77.29
Feb-10	380500	74.01
Mar-10	320500	78.67
Apr-10	368000	85.48
May-10	400500	86.96
Jun-10	407500	70.22
Jul-10	387500	73.74
Aug-10	362000	75.94
Sep-10	410500	76.76
Oct-10	391500	72.26
Nov-10	439500	80.07
Dec-10	422000	91.37

Thus the graph of actual gold price will be:

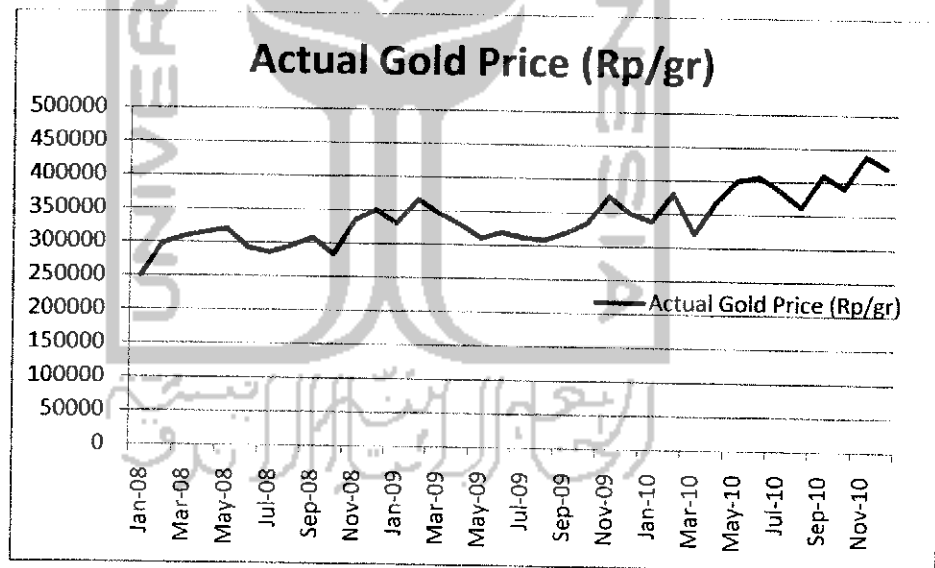


Figure 4.1 Actual Gold graph

and the graph of crude price will be:

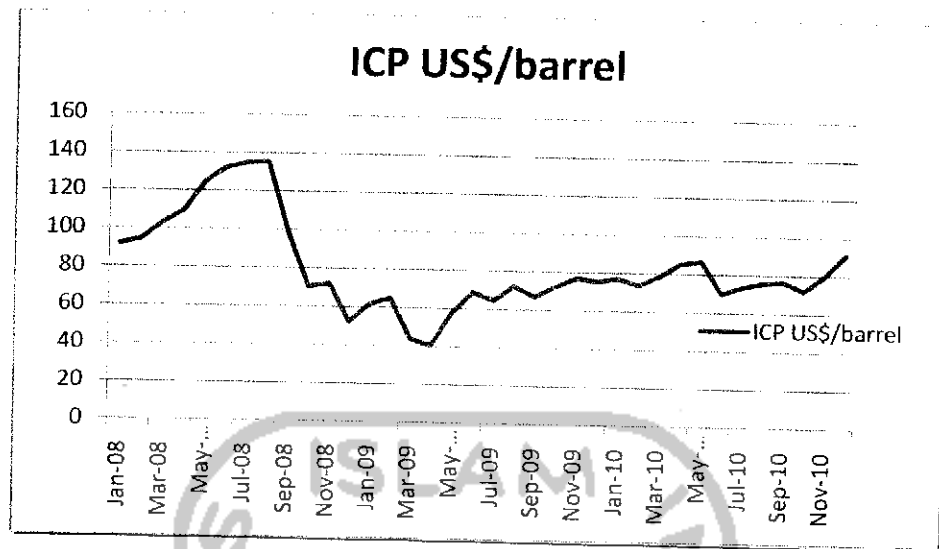


Figure 4.2 Indonesian Crude Price Graph

4.2 Data Processing

4.2.1 Model Selection

Since the relationship between X_k and Y_k are determined by equation 3.1, based on the assumption that the cause of the incident gold price are not observed, resulting that X_k hidden under observation data Y_k . So, a pair of $\{ X_k , Y_k \}$ is Hidden Markov Model. Based on previous chapter, there are three models on Hidden Markov. Second model with discrete time and continuous observation is characterized by several parameters including the transition probability matrix, expectation and variance. The causative factor of gold price is assumed as a state (discrete) that not observed directly and has the properties of Markov chains. Gold price in a specified period is used as data representing a continuous observation. Therefore second model which is Hidden markov model with discrete time and continuous observation is being used. There are 36 observation sequences at January 2008 – December 2010. Y_k are estimated with \hat{Y}_k . This estimated value is expected to accost the actual value.

4.2.2 Algorithm Selection

Estimation is being done using algorithms described in Equation 3.11 to 3.25 in chapter III. Set $N = 2$, $T = 36$ and data input $\{Y_k\}$ as same as data table.

Parameters tested based on the simulation model includes:

1. A Parameter (transition probability matrix) sized 2×2
2. C Parameter shaped vector 2×1
3. σ Parameter shaped vector 2×1

Set the value of A , C and σ :

$$A_{ij} = \begin{pmatrix} 10/20 & 10/20 \\ 9/15 & 6/15 \end{pmatrix} \quad C_1 = \begin{pmatrix} 1400 \\ 1200 \end{pmatrix} \quad \sigma_1 = \begin{pmatrix} 150 \\ 200 \end{pmatrix}$$

$$C_2 = \begin{pmatrix} 2200 \\ 1500 \end{pmatrix} \quad \sigma_2 = \begin{pmatrix} 300 \\ 250 \end{pmatrix}$$

4.2.3 Data Processing

Thus the state of every month described in table 4.4 as follow:

Table 4.4 State in every month

Month	Price (Rp/gr)	State
January 2008	250000	1
February 2008	298000	1
March 2008	308500	1
April 2008	315000	1
May 2008	320000	1
June 2008	292000	2
July 2008	285500	2
August 2008	295500	1
September 2008	308500	1
October 2008	283500	2
November 2008	335000	1
December 2008	350500	1

January 2009	320500	2
February 2009	367000	1
March 2009	346500	2
April 2009	330000	2
May 2009	310000	2
June 2009	315000	1
July 2009	319500	2
August 2009	309000	2
September 2009	319500	1
October 2009	354500	1
November 2009	385500	1
December 2009	350000	2
January 2010	338500	2
February 2010	380500	1
March 2010	320500	2
April 2010	368000	1
May 2010	400500	1
June 2010	397500	2
July 2010	347500	2
August 2010	362000	1
September 2010	410500	1
October 2010	391500	2
November 2010	439500	1
December 2010	422000	2

And the state transition described in table 4.5 as follow:

Month	Price (Rp/gr)	State	Transition
January 2008	250000	1	
February 2008	298000	1	1-1
March 2008	308500	1	1-1
April 2008	315000	1	1-1
May 2008	320000	1	1-1
June 2008	292000	2	1-2
July 2008	285500	2	2-2
August 2008	295500	1	2-1
September 2008	308500	1	1-1
October 2008	283500	2	1-2
November 2008	335000	1	2-1
December 2008	350500	1	1-1

January 2009	320500	2	1-2
February 2009	367000	1	2-1
March 2009	346500	2	1-2
April 2009	330000	2	2-2
May 2009	310000	2	2-2
June 2009	315000	1	2-1
July 2009	311500	2	1-2
August 2009	309000	2	2-2
September 2009	319500	1	2-1
October 2009	354500	1	1-1
November 2009	385500	1	1-1
December 2009	350000	2	1-2
January 2010	338500	2	2-2
February 2010	380500	1	2-1
March 2010	320500	2	1-2
April 2010	368000	1	2-1
May 2010	400500	1	1-1
June 2010	397500	2	1-2
July 2010	347500	2	2-2
August 2010	362000	1	2-1
September 2010	410500	1	1-1
October 2010	391500	2	1-2
November 2010	439500	1	2-1
December 2010	422000	2	1-2

In the same algorithm on chapter III, the trials for predetermined parameter was conducted using Matlab® and the results are presented in table 4.6 :

a. Simulation of $A - C_t - \sigma_t$

Table 4. 6 Simulation Result of $A - C_t - \sigma_t$

Month	Actual Price (Rp/gr)	Simulation Price (Rp/gr)	Galat (%)
January 2008	250000	235500	5.8
February 2008	298000	280000	6.4
March 2008	308500	297000	3.6
April 2008	315000	299500	4.9
May 2008	320000	329000	-2.8

June 2008	292000	290500	0.5
July 2008	285500	287000	-0.6
August 2008	295500	280500	5.1
September 2008	308500	300500	2.5
October 2008	283500	269000	5.2
November 2008	335000	310000	7.4
December 2008	350500	345500	1.6
January 2009	320500	335000	-4.4
February 2009	367000	360500	1.6
March 2009	346500	320000	7.6
April 2009	330000	345000	-4.1
May 2009	310000	319000	-3.5
June 2009	315000	323500	-2.7
July 2009	311500	321500	-3.6
August 2009	309000	310000	-0.5
September 2009	319500	325500	-2.0
October 2009	354500	375000	-6.0
November 2009	385500	380000	1.3
December 2009	350000	361500	-2.5
January 2010	338500	322000	4.9
February 2010	380500	361000	5.2
March 2010	320500	346000	-7.9
April 2010	368000	369500	-0.5
May 2010	400500	387000	3.5
June 2010	397500	370500	6.7
July 2010	347500	358000	-2.6
August 2010	362000	368000	-1.8
September 2010	410500	399500	2.7
October 2010	391500	401000	-2.4
November 2010	439500	420500	4.4
December 2010	422000	415000	1.8

Total percentage error of this simulation is 2.25%. Thus the comparison graph between actual and $A - C_I - \sigma_I$ simulation will be described in figure 4.3 :

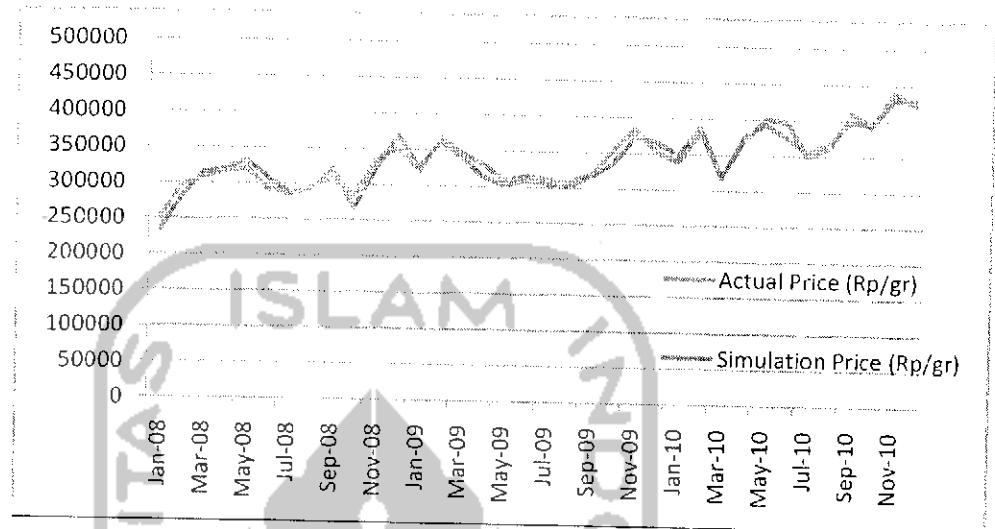


Figure 4.3 Actual vs Simulation of $A - C_1 - \sigma_1$

b. Simulation of $A - C_1 - \sigma_2$

Table 4. 7 Simulation Result of $A - C_1 - \sigma_2$

Month	Actual Price (Rp/gr)	Simulation Price (Rp/gr)	Galat (%)
January 2008	250000	245000	2.0
February 2008	298000	287500	3.4
March 2008	308500	285000	7.5
April 2008	315000	321000	-1.9
May 2008	320000	330000	-3.1
June 2008	292000	288500	1.2
July 2008	285500	280500	1.8
August 2008	295500	292500	1.0
September 2008	308500	311500	-1.0
October 2008	283500	274000	3.1
November 2008	335000	315500	5.8
December 2008	350500	340000	3.1
January 2009	320500	340000	-6.2
February 2009	367000	361500	1.3
March 2009	346500	321000	7.3
April 2009	330000	319000	3.6
May 2009	310000	320000	-4.0
June 2009	315000	321500	-2.1
July 2009	311500	316000	-1.9
August 2009	309000	305000	1.3

September 2009	319500	338000	-5.9
October 2009	354500	365000	-3.2
November 2009	385500	370000	3.7
December 2009	350000	360500	-2.3
January 2010	338500	332000	1.9
February 2010	380500	374500	1.6
March 2010	320500	318000	0.9
April 2010	368000	373500	-1.6
May 2010	400500	405500	-1.1
June 2010	397500	370500	6.7
July 2010	347500	375000	-7.6
August 2010	362000	361000	0.5
September 2010	410500	399500	2.7
October 2010	391500	381000	2.8
November 2010	439500	422500	3.9
December 2010	422000	430000	-2.0

Total percentage error of this simulation is 2.22%. Thus the comparison graph between actual and simulation of $A - C_2 - \sigma_2$ described in figure 4.4:

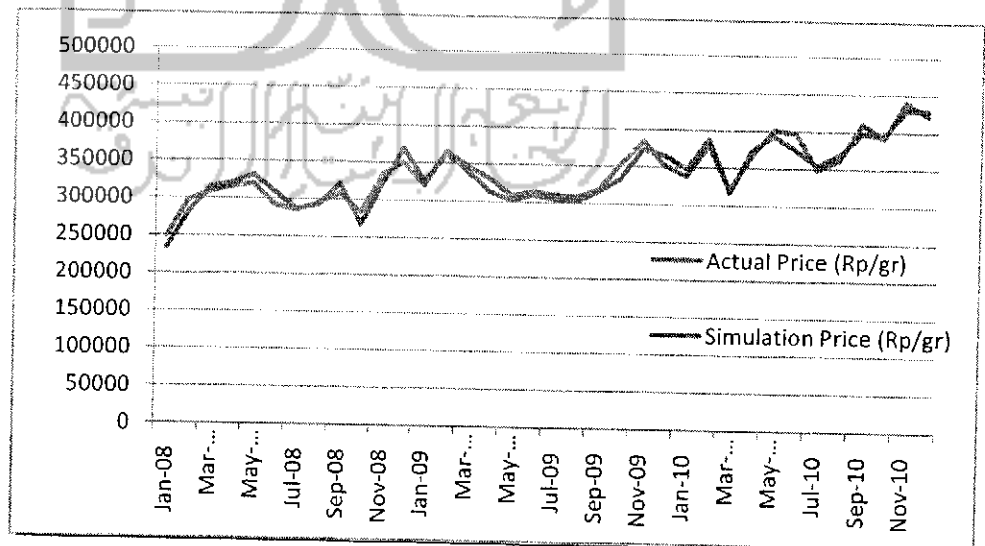


Figure 4.4 Actual vs Simulation of $A - C_1 - \sigma_2$

c. Simulation of $A - C_2 - \sigma_1$ Table 4. 8 Simulation Result of $A - C_2 - \sigma_1$

Month	Actual Price (Rp/gr)	Simulation Price (Rp/gr)	Galat (%)
Jan-08	250000	269500	-7.8
Feb-08	298000	290000	3.2
Mar-08	308500	292500	5.2
Apr-08	315000	312000	0.9
May-08	320000	334000	-4.4
Jun-08	292000	288500	1.2
Jul-08	285500	292000	-2.4
Aug-08	295500	283500	4.0
Sep-08	308500	320500	-3.9
Oct-08	283500	265000	6.6
Nov-08	335000	319500	4.6
Dec-08	350500	364500	-3.8
Jan-09	320500	334000	-4.3
Feb-09	367000	372500	-1.7
Mar-09	346500	330500	4.6
Apr-09	330000	341000	-3.0
May-09	310000	314000	-1.8
Jun-09	315000	323500	-2.7
Jul-09	311500	318500	-2.6
Aug-09	309000	311000	-0.6
Sep-09	319500	321500	-0.7
Oct-09	354500	369000	-4.3
Nov-09	385500	375000	2.4
Dec-09	350000	350500	0.5
Jan-10	338500	331000	2.3
Feb-10	380500	388500	-2.1
Mar-10	320500	318500	0.6
Apr-10	368000	348500	5.2
May-10	400500	407500	-1.6
Jun-10	397500	384500	3.2
Jul-10	347500	352000	-1.0
Aug-10	362000	373000	-3.0
Sep-10	410500	399500	2.7
Oct-10	391500	389500	0.6

Nov-10	439500	417500	5.1
Dec-10	422000	413500	2.1

Total percentage error of this simulation is 2.96%. Thus the comparison graph between actual and simulation of $A - C_1 - \sigma_2$ described in figure 4.5:

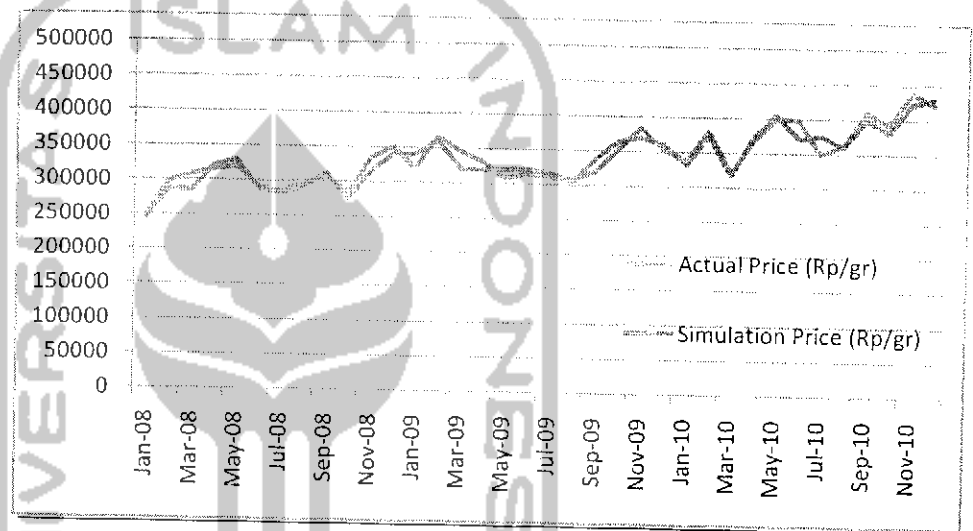


Figure 4.5 Actual vs Simulation of $A - C_2 - \sigma_1$

d. Simulation of $A - C_2 - \sigma_2$

Table 4.9 Simulation Result of $A - C_2 - \sigma_2$

Month	Actual Price (Rp/gr)	Simulation Price (Rp/gr)	Galat (%)
Jan-08	250000	234500	6.7
Feb-08	298000	281500	5.5
Mar-08	308500	315000	-2.0
Apr-08	315000	320500	-1.7
May-08	320000	331500	-3.5
Jun-08	292000	308500	-5.6
Jul-08	285500	287000	-0.6
Aug-08	295500	292500	1.1
Sep-08	308500	321500	-4.2
Oct-08	283500	266500	6.0
Nov-08	335000	323000	3.6

Dec-08	350500	369500	-5.3
Jan-09	320500	325000	-1.5
Feb-09	367000	362500	1.0
Mar-09	346500	339000	2.1
Apr-09	330000	313500	5.3
May-09	310000	302500	2.0
Jun-09	315000	309500	1.8
Jul-09	311500	305000	1.8
Aug-09	309000	302000	2.1
Sep-09	319500	316500	0.9
Oct-09	354500	334000	5.6
Nov-09	385500	373500	2.9
Dec-09	350000	365500	-3.6
Jan-10	338500	348000	-2.8
Feb-10	380500	388000	-2.0
Mar-10	320500	317500	0.9
Apr-10	368000	374500	-1.8
May-10	400500	393000	2.0
Jun-10	397500	374500	5.7
Jul-10	347500	354000	-1.5
Aug-10	362000	369000	-1.9
Sep-10	410500	397500	3.2
Oct-10	391500	394000	-0.7
Nov-10	439500	429500	2.3
Dec-10	422000	429000	-1.6

Total percentage error of this simulation is 2.8%. Thus the comparison graph between actual and simulation of $A - C_2 - \sigma_2$ described in figure 4.6:

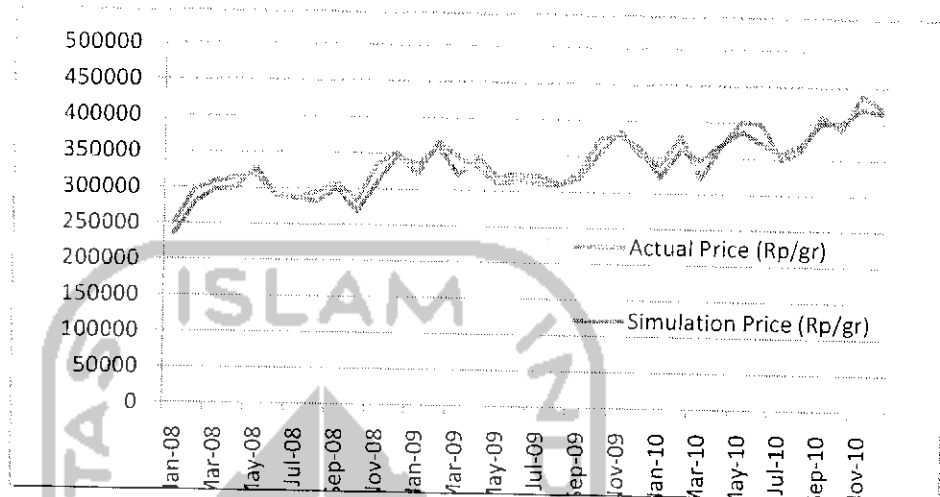


Figure 4.6 Actual vs Simulation of $A - C_2 - \sigma_2$

UNIVERSITAS ISLAM INDONESIA
 الجامعة الإسلامية
 الرابطة الإسلامية العالمية

CHAPTER V

DISCUSSION

Chapter V is the discussion about the result obtain in chapter IV.

According to chapter I, the main issue is about Hidden Markov in time series forecasting model. The principle of HMM is that the current price influenced by $n-1$ prices. Gold price is a stochastic process and its causative factor formed a markov chain. New academic year and crude oil selected to be causative factors in this case. In the new academic year in June-July, the data suggest that gold price were inceasing in June 2009 and June 2010 but decreasing in July or one month thereafter. This condition is due to the community needs for school purposes. The effect of crude oil at the gold price occured in September 2008, October 2008, December 2008, March 2009, April 2009, July 2009, December 2009, February 2010, June 2010, and October 2010 where the decreasing in crude oil contributed to the gold prices in that month or a month thereafter.

π is the initial probability. Define π for state 1 = 0.573 and π for state 2 = 0.427. C and σ is the expectation and variance of data. To facilitate the search for the parameter, created a program functional programming-based computing using Matlab®. The estimators obtained were used to calculate the expected value of gold price.

With the average percentage error of 2.25% for $A - C1 - \sigma1$, 2.22% for $A - C1 - \sigma2$, 2.96% for $A - C2 - \sigma1$, and 2.8% for $A - C2 - \sigma2$ which is less than 10%, then Hidden Markov Model is sufficient to be implement on gold price by means of two causative factors. With the same value of c and enlarged σ , the error can be narrowed. Reciprocally

by the same values of σ and c that are enlarged, the error can also be narrowed. It proves that simulation sequence is closed to the actual one.



CHAPTER VI

CONCLUSION AND SUGGESTION

6.1 Conclusion

According to the explanation in chapter V, some conclusions can be established is as follow:

1. Market trend of gold in 2008 was decreasing on June, July, August and October. While in the 2009 decreasing point was on January, March, May, July, August and December. For the 2010 decreasing point was on January, March, June, July and December based on the Simulation of $A - C1 - \sigma1$. Market trend of gold in 2008 was decreasing on March, June, July and October. While in the 2009 decreasing point was on March, April, July, August and December. In 2010, decreasing point was on January, March, June, August and October based on the Simulation of $A - C1 - \sigma2$. The gold trend for the simulation of $A - C2 - \sigma1$ in 2008 was decreasing on June, August and October. While in 2009, the decreasing point was on January, March, May, July, August and December. In 2010, decreasing point was on January, March, June, July, October and December. The gold trend for the simulation of $A - C2 - \sigma2$ in 2008 was decreasing on June, July and October. While in 2009, the decreasing point was on January, March, April, May, July, August and December. For 2010 decreasing point was on January, March, June, July and October.

6.2 Suggestion

The suggestion for further research can be described as follow:

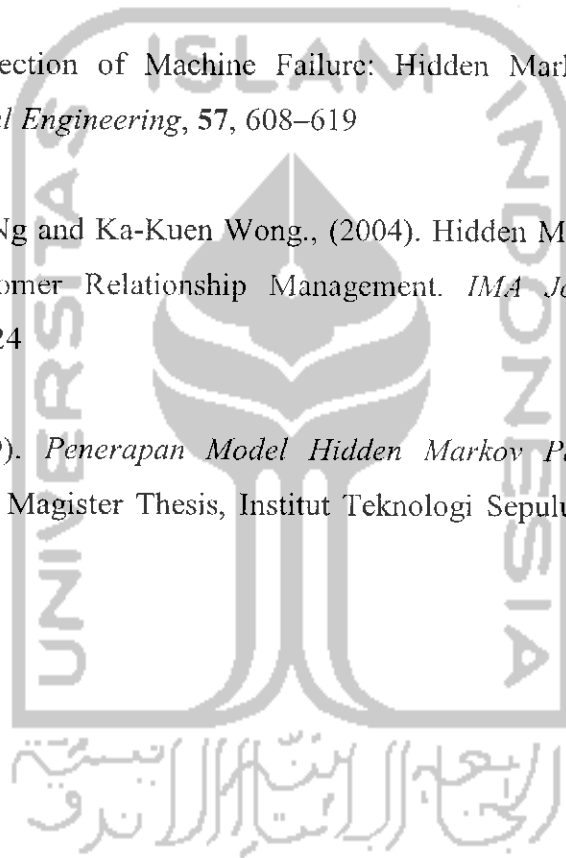
1. In the case of stochastic situation, other method need to be added to minimize the error.
2. Further research about Hidden Markov model needs to be conducted involving 3 or more number of factors on the different field of stochastic situations.



REFERENCES

- Li M., (2000). *Probabilistic System Analyses: Part I*. University of Witwatersrand, Johannesburg
- Nuri et al., (2008). Analisa Teknik Pengenalan Sinyal Wicara Dengan Hidden Markov Model-Neural Network. *Jurnal litek* , vol 5, no 2, 64-67
- Rabiner, L.R., Juang, B.H. (1986), *Hidden Markov Model*, Springer-Verlag. New York.
- Pratiwi, A., (2006). Algoritma Viterbi dalam Metode Hidden Markov Models pada Teknologi Speech Recognition. *Laboratorium Ilmu dan Rekayasa Komputasi* , Departemen Teknik Informatika, Institut Teknologi Bandung
- Pratiwi, A., (2003). *On-Line Chinese Character Recognition Using Hidden Markov Models*. Master Theses Informatics Engineering, Institut Teknologi Sepuluh Nopember Surabaya. (Unpublished)
- Rabiner, L.R., Wheelwright, S.C., McGee. (1999), *Metode dan Aplikasi Peramalan*, Binarupa Aksara, Jakarta.
- Ad. Rafiul Hassan and Baikunth Nath., (2005). StockMarket Forecasting Using Hidden Markov Model: A New Approach. *Proceedings of the 2005 5th International Conference on Intelligent Systems Design and Applications (ISDA'05)*
- Muhammad Imam., (2007). *An Application of Face Recognition on Digital Image using Embedded Hidden Markov Models*. Thesis, Institut Pertanian Bogor. (Unpublished)
- Nur Fathoni., (2008) . *Kajian Hidden Markov Kontinyu dan Aplikasinya pada Harga Gabah Kering Panen*. Mathematic Magister Theses, Institut Pertanian Bogor. (Unpublished)

- oiner, L.R., Wilpon J.G., (1987). Application of Hidden Markov Model to Automatic Speech Endpoint Detection. *Computer Speech and Language*, vol 2, hal 321-341
- riawaty, Kristina. (2005). Pendugaan Parameter Model Hidden Markov. *Jurnal Matematika dan Aplikasinya*. vol. 4
- ri, Allen H. (2008). Detection of Machine Failure: Hidden Markov Approach. *Journal Computers & Industrial Engineering*, 57, 608–619
- ai-Ki Ching, Michael K. Ng and Ka-Kuen Wong., (2004). Hidden Markov Models and Their Applications to Customer Relationship Management. *IMA Journal of Management Mathematics* , 15, 13–24
- uni Dewi Lestari., (2009). *Penerapan Model Hidden Markov Pada Peramalan Harga Premium*. Mathematic Magister Thesis, Institut Teknologi Sepuluh Nopember Surabaya. (Unpublished)



ATTACHMENT

```
clc;
clear;
%initial parameter
N=2 %number of state
A=[10/20 10/20; 9/15 6/15]
%A=matrix transition probability
phi=phi(A)
c=[1400; 1200]
%c2 [2200; 1500]
sigma=[150; 200]
%sigma=[300; 250]
%data
logammulia
%normalisasi data
n0=norm(x);
v=x/n0;
a0=1;
b0=40;
v=a0-v*(a0-b0);
y=v;
n=length(y);

%lakukan i 1 sampai n (banyaknya data)
for i =1:n
    i;
    %inisialisasi parameter
    pendugaX=phi1;
    pendugaLomp=zeros(N,N);
    pendugaWaktu=zeros(N,1);
    pendugaProses=zeros(N,1);
    pendugaProses2=zeros(N,1);

    for j=0:(i-1)
        j;
        %penduga rekursif
        %penduga state
        pendugax=zeros(N,1);
        for k=1:N
            a= pendugaX'*gama(k, y(j+1), c, sigma, N)*A(:,k);
            pendugax= pendugax + a;
        end
        pendugax;

        %penduga lompatan
        pendugalomp=zeros(N,1);
        for k=1:N
            a= pendugaLomp*gama(k, y(j+1), c, sigma, N).*A(:,k);
            pendugalomp= pendugalomp + a;
        end

        for r=1:N
            for s=1:N
                B= pendugaX'*gama(r, y(j+1), c, sigma, N)*A(s,r)*unit(N,s);
                if r==1
                    B= pendugalomp(1)+B;
                    B1(:,s,:)=B;
                else

```

```

        B=pendugalomp(2)+B;
        B2(:,s,:)=B;
    end
end
end
B3=[B1;B2];
% <pendugalompatan,1>
B=[sum(B1); sum(B2)];

%penduga waktu
pendugawaktu=zeros(N,1);
for k=1:N
    a= pendugaWaktu'*gama(k, y(j+1), c, sigma, N)*A(:,k);
    pendugawaktu= pendugawaktu + a;
end

for r=1:N
    C= pendugaX'*gama(r, y(j+1), c, sigma, N)*A(:,r) +
pendugawaktu(r);
    C1(:,r,:)=C;
end
C2=C1';
% <pendugawaktu,1>
C=sum(C2,2) ;

%penduga proses
pendugaproses=zeros(N,1);
for k=1:N
    a= pendugaProses'*gama(k, y(j+1), c, sigma, N)*A(:,k);
    pendugaproses= pendugaproses + a;
end

for r=1:N
    D= pendugaX'*gama(r, y(j+1), c, sigma, N)*A(:,r)*y(j+1) +
pendugaproses(r);
    D1(:,r,:)=D;
end
D2=D1';
% <pendugaproses,1>
D=sum(D2,2) ;

%penduga proses2
pendugaproses2=zeros(N,1);
for k=1:N
    a= pendugaProses2'*gama(k, y(j+1), c, sigma, N)*A(:,k);
    pendugaproses2= pendugaproses2 + a;
end

for r=1:N
    E= pendugaX'*gama(r, y(j+1), c, sigma, N)*A(:,r)*y(i)^2 +
pendugaproses2(r);
    E1(:,r,:)=E;
end
E2=E1';
% <pendugaproses,1>
E=sum(E2,2) ;

%penduga
asr =[ B(1,:)./C(1) ;B(2,:)./C(2)];
cr= C./D;

```

```

sig=(E-2*cr.^D/cr.^2.*C)./C;
cs=cr + (sig.^(1/2));
cs2(:,i,:)=cs;

%mendapat phibaru
phibaru=phi(asr);
phi2(:,i,:)=phibaru;

%perbarui buat looping
pendugaX=pendugax;
pendugaLomp=B;
pendugaWaktu=C;
pendugaProses=D;
pendugaProses2=E;
end
end

[phi2' cs2'];
Y=cs2.*phi2;
Y=sum(Y);
Y;
[ y' Y'];

%denormalisasi data
Y1=(a0-Y)/(a0-b0);
Y=Y1 * n0;

plot(1:n-1, x(2:n), '-r', 1:n, Y, '-b')
set(gcf,'color','w')
xlabel('Data ke-', 'color','m', 'fontweight','bold')
ylabel('Nilai Data', 'color','m', 'fontweight','bold')
titstrl=('Plot Data Aktual vs Data Ramalan pake Hidden Markov');
title(titstrl, 'color','b', 'fontweight','bold');
grid on
legend('Data Aktual', 'Data Ramalan')
%text(25,5500,['R2 =', num2str(R2)], 'color','k', 'fontweight','bold', 'fontsize',
',13);

a=x(2:n);
b=Y(1:n-1);
e=abs(a-b);

fprintf('\n\nHasil \n')
fprintf('\n=====')
fprintf('\n')
fprintf('      Data Aktual      Data Ramalan      Galat      ')
fprintf('\n=====')
[ a' b' e']

%data statistik
R2=1-((sum((a-b).^2))/(sum((a-mean(b)).^2)))

%MPE
MPE=(1/(n-1)*sum((a-b)./b))*100

%MAPE
MAPE=(1/(n-1)*sum(abs((a-b)./b)))*100
MSE= sum((a-b).^2)/(n-1)

```



ramalan=Y(n)

N =

2

A =

0.5

0.5

0.6

0.4

phi =

0.573

0.427

c =

1400

1200

sigma =

150

200

Hasil

Data Aktual	Data Ramalan	Galat
250000	235500.000000000	14500
297708.2	278664.649527242	19043.6
308445.06	297201.946728513	11243.1
315012.22	299504.443102659	15507.8
320056.4	329065.316225945	9008.916
291885.21	290519.546457540	1365.66
285471.08	287056.116452002	1585.036
295466.09	280520.654662057	14945.4
308338.2	300501.131165948	7837.07
283488.35	268865.000489726	14623.3
334825	310125.013101043	24700
351021.04	345522.364808085	5498.68
320455.45	334695.250458400	14239.8
366289.2	360482.460035794	5806.74
346375.11	320194.644308750	26180.5
331102.4	344812.344209678	13709.94

308500.46	319200.803397864	10700.34
315000	323505.719658220	8505.72
310224.16	321296.157666028	11072
308821.82	310248.003167955	1426.183
319270	325514.602259734	6244.602
353789.2	374992.146830670	21202.95
384259.3	379306.280342454	4953.02
352441.1	361339.720067140	8898.62
338624.15	322176.346220580	16447.8
380502.8	360775.169735006	19727.6
320442.54	345830.643195200	25388.1
367650.4	369514.373716455	1863.974
401008.96	387137.861340069	13871.1
397025.6	370509.773568421	26515.8
348612.4	357691.160875564	9078.761
362312.65	368942.816490056	6630.166
410449.14	399377.073814525	11072.1
391605	401168.570339452	9563.57
439738.11	420485.800034158	19252.3
422250.24	414594.159618660	7656.08

R2 = 0.437568943742

MAPE = 3.6153772478246

ramalan = 373056.42136793

N = 2

A = 0.5 0.5
0.6 0.4

phi = 0.573
0.427

c =

1400

1200

sigma =

300

250

Hasil

Data Aktual	Data Ramalan	Galat
250000	245089.717202674	-4910.28
297708.2	287520.418763050	-10187.8
308445.06	299200.604504300	-9244.46
315012.22	321054.044025820	6041.824
320056.4	329886.044068077	9829.644
291885.21	288525.403151406	-3359.81
285471.08	280445.780354450	-5025.3
295466.09	292369.405448000	-3096.68
308338.2	311511.046176096	3172.846
283488.35	274775.147000000	-8713.2
334825	315512.710310013	-19312.3
351021.04	340196.470340300	-10824.6
320455.45	340201.174304000	19745.72
366289.2	361511.971510454	-4777.23
346375.11	321088.464780412	-25286.6
331102.4	339155.302967153	8052.903
308500.46	313925.009556104	5424.55
315000	321521.065971264	6521.066
310224.16	316145.946738250	5921.787
308821.82	304865.201958621	-3956.62
319270	338243.448501245	18973.45
353789.2	365064.479120234	11275.28
384259.3	377165.840165494	-7093.46
352441.1	360522.549745230	8081.45
338624.15	332204.589256410	-6419.56
380502.8	374525.154620548	-5977.65
320442.54	317654.513249710	-2788.03
367650.4	373368.145404360	5717.745
401008.96	405298.801125821	4289.841
397025.6	380501.164101546	-16524.4

348612.4	335221.761002115	-13390.6
362312.65	360664.495322106	-1648.15
410449.14	399513.864441025	-10935.3
391605	380680.344059700	-10924.7
439738.11	422517.196546208	-17220.9
422250.24	410795.830049615	-11454.4

R2 =
0.647220945313

MAPE =
2.650225678423

ramalan =
373048.468216069

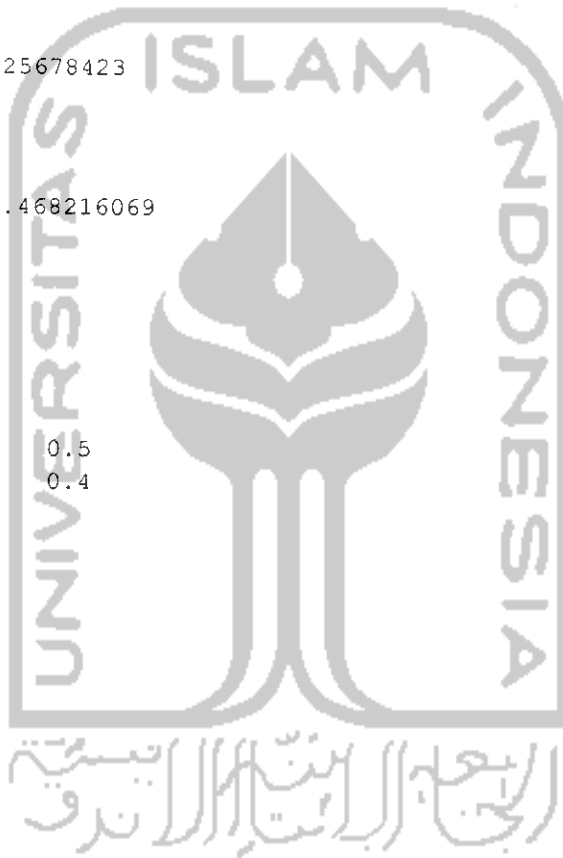
N =
2

A =
0.5 0.5
0.6 0.4

phi =
0.573
0.427

c =
2200
1500

sigma =
150
200



Hasil

Data Aktual	Data Ramalan	Galat
250000	269500.000000000	19500
297708.2	288245.499125672	-9462.7

308445.06	292350.601244952	-16094.5
315012.22	312156.884316250	-2855.34
320056.4	334198.900364515	14142.5
291885.21	288365.779201990	-3519.43
285471.08	292212.986123450	6741.906
295466.09	283520.123106441	-11946
308338.2	320465.805990412	12127.61
283488.35	264857.300964552	-18631
334825	319499.031212956	-15326
351021.04	364502.116534629	13481.08
320455.45	334105.941202564	13650.49
366289.2	372400.574366315	6111.374
346375.11	330511.416008304	-15863.7
331102.4	341198.465401233	10096.07
308500.46	314095.446215380	5594.986
315000	323560.495110685	8560.495
310224.16	318336.306248812	8112.146
308821.82	310789.645236904	1967.825
319270	321521.642199535	2251.642
353789.2	369102.226401583	15313.03
384259.3	375165.703660904	-9093.6
352441.1	350521.165990053	-1919.93
338624.15	330773.466052311	-7850.68
380502.8	388467.610003020	7964.81
320442.54	318522.996458040	-1919.54
367650.4	348515.334658016	-19135.1
401008.96	407506.095349722	6497.135
397025.6	384369.641580749	-12656
348612.4	352180.347025911	3567.947
362312.65	373096.820640553	10784.17
410449.14	399520.518940564	-10928.6
391605	389350.108920528	-2254.89
439738.11	417445.528077530	-22292.6
422250.24	413511.540463446	-8738.7

R2 = 0.573258227024

MAPE = 2.962758053751

ramalan =
372845.63214611

N =
2

A =
0.5 0.5
0.6 0.4

phi =
0.573
0.427

c =
2200
1500

sigma =
300
250

Hasil

Data Aktual	Data Ramalan	Galat
250000	233251.486912531	-16748.513
297708.2	281325.632184598	-16382.568
308445.06	314582.642795310	6137.5828
315012.22	320289.946821549	5277.72682
320056.4	331332.846791135	11276.4468
291885.21	308312.546445127	16427.3364
285471.08	287098.647825910	1627.56783
295466.09	292290.049563020	-3176.0404
308338.2	321405.849651008	13067.6497
283488.35	266415.287901653	-17073.062
334825	322921.027944562	-11903.972
351021.04	369502.577910465	18481.5379
320455.45	325104.959621470	4649.50962
366289.2	362448.589045618	-3840.611
346375.11	339115.995482630	-7259.1145



331102.4	313465.075889134	-17637.324
308500.46	302294.654774211	-6205.8052
315000	309335.707146689	-5664.2929
310224.16	304586.558749026	-5637.6013
308821.82	302205.056443791	-6616.7636
319270	316300.544301970	-2969.4557
353789.2	334112.640227154	-19676.56
384259.3	373287.446915700	-10971.853
352441.1	365278.114908735	12837.0149
338624.15	348187.461797332	9563.3118
380502.8	388055.462080543	7552.66208
320442.54	317486.877034915	-2955.663
367650.4	374245.479025193	6595.07903
401008.96	392841.614730595	-8167.3453
397025.6	374510.477102921	-22515.123
348612.4	353765.110295348	5152.7103
362312.65	369200.227109255	6887.57711
410449.14	397408.144100695	-13040.996
391605	394240.479256301	2635.47926
439738.11	429558.479254460	-10179.631
422250.24	429005.695530253	6755.45553

R2 = 0.497326245215

MAPE = 2.812467393757

ramalan = 372651.75443784