FORECASTING WITH IRREGULAR DEMAND HISTORY

Athapol Ruangkanjanases
Doctoral Candidate
Stuart School of Business
Illinois Institute of Technology
565 West Adams Street
Chicago, IL 60661
athapol@stuart.iit.edu

Nick T. Thomopoulos, Ph.D.
Professor of Management Science
Stuart School of Business
Illinois Institute of Technology
565 West Adams Street
Chicago, IL 60661
thomop@stuart.iit.edu

ABSTRACT

Occasionally, demand data are not issued in equal time periods and even some demand histories are missing. This paper presents a forecasting technique which enables the generation of forecasts when demand data are not issued in the uniform time space associated with missing demand history. The coefficient of variation of the forecast errors (COV) is used to measure the performance of forecasting. The relationships between the percentage of missing data, the number of observations and the COV are also discussed.

1. INTRODUCTION

Data collections are usually issued in equally segregated length of time such as weekly, monthly, quarterly or yearly basis. However, there are some occasions where items are inexpensive and these items need not be registered every time they are in demand. The demands are recorded occasionally rather than periodically. In some cases, missing data or stock-outs even occur before the demand for the next time period can be recorded. Therefore, the forecasters deal with two problems: the missing demand history and the unequal intervals in time series. This paper is part of the dissertation of Athapol Ruangkanjanases for the Ph.D. in Management Science at the Stuart School of Business, Illinois Institute of Technology.

A number of studies have been done regarding the forecasting with unequal time intervals or with missing data. Brown (1963) mentions the need for forecasting where demands are implicit over time periods. He gives an example of an inventory situation where there are no tight records or controls. Therefore, a forecaster may find it difficult to get the actual periodical demands. Brown shows that those items can still be forecasted by using a simple interpolation to estimate the demand rate. The order point is found by extrapolation.

Dunsimuir and Robinson (1981) show how to use the autoregressive moving average (ARMA) in the cases where a number of data are missing. Aldrin and Damsleth (1989) propose another unified approach to handling missing data within the smoothing methods. It is compared with previously suggested modifications using 12 non-seasonal time series taken from the literature. The result shows that the smoothing methods, properly modified, usually perform well if the time series have a moderate number of missing observations.

Manaspiti (1990) shows two techniques dealing with the situation where demands cannot be nominated in equal lengths of time. The first formulation is concerned with the horizontal model and the methodology is based on the exponential smoothing and the moving average. The second technique copes with the trend model. The approach is developed from the conventional linear regression model.

1.1 Causes of Problem

The following are examples of situations where the data may not be issued in a uniform time space and some data are missing.

1.1.1 ***Unorganized Data Collection.*** For instance, one data entry is the demand for April and first five days of May, while the next entry is for the last 10 days of May plus the entire month of June. In this case, demand for the month of May is part of both entries. In addition, there are some missing data in the month of May.

1.1.2 ***Nature of Occurrences.*** At some point of time the data may not be available or retrievable. For example, the Severe Acute Respiratory Syndrome (SARS) Virus cases were reported sparsely when they were first discovered in Asia in 2002. There was no intention to keep the record every period since it had not become routine then.

1.1.3 ***Delayed Deliveries and Unexpected Increases in Demand.*** Delays can occur because of weather conditions, deliveries of wrong materials, quality problems and so on. Unexpected increases in demand are also the cause of stock-out.

1.2 Objectives

1) To propose techniques that enable the generation of forecasts when data are not issued in the uniform time space associated with missing demand history. The Trend Model is proposed in this paper. More techniques for other models such as Horizontal Model, Three-Term Model (Seasonal Model), and Four-Term Model (Trend Seasonal Model) can be found in Ruangkanjanases (2007).

2) To investigate the performance of forecasting measured by the value of coefficient of variation of the forecast errors (COV). The idea is to explore how the number of observations and the percentage of missing data have an effect on the performance of forecasting.

2. FORECASTING MODEL

Trend Model

The trend model copes with the situation when the change of demand level has a certain relation with time. In other words, the rate change of demand levels with respect to time is relatively constant.

Let

μ_t = demand level at time t

a = intercept or the level at time t= 0

b = slope or the rate of increase or decrease of demand

The demand level is a function of t, i.e.,

$$\mu_t = a + bt \tag{2.1}$$

and

$D_i =$ total demand from time t_{1i} to t_{2i}

Where t_{1i} = time at starting point of observation i

t_{2i} = time at ending point of observation i

$\hat{D}_i$ = estimate of D_i

Since

$$D_i = \int_{t_{1i}}^{t_{2i}} \mu_t \, dt$$

then

$$\hat{D}_i = \int_{t_{1i}}^{t_{2i}} \hat{\mu}_t \, dt = \int_{t_{1i}}^{t_{2i}} (\hat{a} + \hat{b}t) dt$$

Therefore,

$$\hat{D}_i = \hat{a}(t_{2i} - t_{1i}) + \frac{\hat{b}}{2}(t_{2i}{}^2 - t_{1i}{}^2) \qquad (2.2)$$

Where $\hat{a}$ = estimate of a

$\hat{b}$ = estimate of b

For instance, the demands are listed for the past k intervals and some demand data are missing between intervals. The estimate total demand from observations 1 to k can be calculated as shown in Table 1.

Table 1. Trend Model: calculation of estimate of total demand $\hat{D}_i$, when i = 1 to k

Observation	Starting point of observation i	Ending point of observation i	Total actual demand of observation i	Estimate of D_i
i	t_{1i}	t_{2i}	D_i	$\hat{D}_i = \hat{a}(t_{2i} - t_{1i}) + \frac{\hat{b}}{2}(t_{2i}{}^2 - t_{1i}{}^2)$
1	t_{11}	t_{21}	D_1	$\hat{D}_1 = \hat{a}(t_{21} - t_{11}) + \frac{\hat{b}}{2}(t_{21}{}^2 - t_{11}{}^2)$
2	t_{12}	t_{22}	D_2	$\hat{D}_2 = \hat{a}(t_{22} - t_{12}) + \frac{\hat{b}}{2}(t_{22}{}^2 - t_{12}{}^2)$
.	.	.	.	.
k	t_{1k}	t_{2k}	D_k	$\hat{D}_k = \hat{a}(t_{2k} - t_{1k}) + \frac{\hat{b}}{2}(t_{2k}{}^2 - t_{1k}{}^2)$

Then we can write the equation in matrix form to solve for $\hat{a}$ and $\hat{b}$.

$$\sum_{i=1}^{k} D_i = \sum_{i=1}^{k} \hat{D}_i + errors$$

In other words,

$$\sum_{i=1}^{k} D_i = \sum_{i=1}^{k} \hat{D}_i + \sum_{i=1}^{k} u_i$$

Which means that

$$\begin{bmatrix} D_1 \\ D_2 \\ . \\ . \\ D_k \end{bmatrix} = \begin{bmatrix} t_{21} - t_{11} & \frac{1}{2}(t_{21}{}^2 - t_{11}{}^2) \\ t_{22} - t_{12} & \frac{1}{2}(t_{22}{}^2 - t_{12}{}^2) \\ . & . \\ . & . \\ t_{2k} - t_{1k} & \frac{1}{2}(t_{2k}{}^2 - t_{1k}{}^2) \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_k \end{bmatrix}$$

Using Ordinary Least Square Method (OLS) estimation, the goal is to minimize $\sum_{i=1}^{k} u_i{}^2$ where $\sum_{i=1}^{k} u_i{}^2$ is the sum of the squared errors. In short, we can solve for $\hat{a}$ and $\hat{b}$ by using the matrix to solve the normal equation.

Let $$[X]\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = [Y] \tag{2.3}$$

To find matrix $\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}$, the first step is to multiply the transpose of matrix $[X]$ to both sides of Equation (2.3). We have

$$[X^T X]\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = [X^T]\ [Y] \tag{2.4}$$

Then multiply the inverse of matrix $[X^T X]$ to both sides of Equation (2.4):

$$[X^T X]^{-1}[X^T X]\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = [X^T X]^{-1}[X^T]\ [Y] \tag{2.5}$$

Finally,

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = [X^T X]^{-1}[X^T]\ [Y] \tag{2.6}$$

If

$$[X] = \begin{bmatrix} t_{21}-t_{11} & \frac{1}{2}(t_{21}{}^2 - t_{11}{}^2) \\ t_{22}-t_{12} & \frac{1}{2}(t_{22}{}^2 - t_{12}{}^2) \\ . & . \\ . & . \\ t_{2k}-t_{1k} & \frac{1}{2}(t_{2k}{}^2 - t_{1k}{}^2) \end{bmatrix} \quad \text{and} \quad [Y] = \begin{bmatrix} D_1 \\ D_2 \\ . \\ . \\ D_k \end{bmatrix}$$

Clearly,

$$\begin{bmatrix} t_{21}-t_{11} & \frac{1}{2}(t_{21}{}^2 - t_{11}{}^2) \\ t_{22}-t_{12} & \frac{1}{2}(t_{22}{}^2 - t_{12}{}^2) \\ . & . \\ . & . \\ t_{2k}-t_{1k} & \frac{1}{2}(t_{2k}{}^2 - t_{1k}{}^2) \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} D_1 \\ D_2 \\ . \\ . \\ D_k \end{bmatrix} \tag{2.7}$$

Based on Equation (2.6), therefore,

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \left[\begin{bmatrix} t_{21}-t_{11} & t_{22}-t_{12} & \ldots\ldots & t_{2k}-t_{1k} \\ \frac{1}{2}(t_{21}^2 - t_{11}^2) & \frac{1}{2}(t_{22}^2 - t_{12}^2) & \ldots\ldots & \frac{1}{2}(t_{2k}^2 - t_{1k}^2) \end{bmatrix} \begin{bmatrix} t_{21}-t_{11} & \frac{1}{2}(t_{21}{}^2 - t_{11}{}^2) \\ t_{22}-t_{12} & \frac{1}{2}(t_{22}{}^2 - t_{12}{}^2) \\ . & . \\ . & . \\ t_{2k}-t_{1k} & \frac{1}{2}(t_{2k}{}^2 - t_{1k}{}^2) \end{bmatrix} \right]^{-1} \times$$

$$\begin{bmatrix} t_{21}-t_{11} & t_{22}-t_{12} & \ldots\ldots & t_{2k}-t_{1k} \\ \frac{1}{2}(t_{21}^2 - t_{11}^2) & \frac{1}{2}(t_{22}^2 - t_{12}^2) & \ldots\ldots & \frac{1}{2}(t_{2k}^2 - t_{1k}^2) \end{bmatrix} \begin{bmatrix} D_1 \\ D_2 \\ . \\ . \\ D_k \end{bmatrix} \tag{2.8}$$

Finally, we have

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \sum (t_{2i} - t_{1i})^2 & \sum (\frac{1}{2}(t_{2i}^2 - t_{1i}^2)(t_{2i} - t_{1i}) \\ \sum (t_{2i} - t_{1i})(\frac{1}{2}(t_{2i}^2 - t_{1i}^2)) & \sum (\frac{1}{2}(t_{2i}^2 - t_{1i}^2))^2 \end{bmatrix}^{-1} \times$$

$$\begin{bmatrix} t_{21} - t_{11} & t_{22} - t_{12} & \ldots\ldots & t_{2k} - t_{1k} \\ \frac{1}{2}(t_{21}^2 - t_{11}^2) & \frac{1}{2}(t_{22}^2 - t_{12}^2) & \ldots\ldots & \frac{1}{2}(t_{2k}^2 - t_{1k}^2) \end{bmatrix} \begin{bmatrix} D_1 \\ D_2 \\ . \\ . \\ D_k \end{bmatrix} \quad (2.9)$$

Since the estimated intercept, $\hat{a}$, and the estimated slope, $\hat{b}$, are solved, the total demand for future period can be calculated by using Equation (2.10).

$$\hat{D}_i = \hat{a} + \hat{b}t \quad (2.10)$$

Illustration
Example 1

This example is given to show how to implement the above forecast model. Table 2 shows the actual demands of a product in the last 24 months.

Table 2. Trend Model: raw data of total demand in the past 24 months

Starting Time	Ending Time	Total Demand of Product (Units)
Jan 1, 2005	Feb 12, 2005	308
May 6, 2005	May 31, 2005	220
Sep 24, 2005	Oct 12, 2005	187
Dec 24, 2005	Apr 30, 2006	1530
Jul 1, 2006	Aug 18, 2006	725
Nov 12, 2006	Dec 31, 2006	783

Step I

Convert the raw data of starting time and ending time of each observation to numerical data. The data of converted observation time is shown in Table 3. The data from Table 3 are then plotted in the chart in Figure1.

Table 3. Trend Model: total demand in the past twenty four months

Observation i	Starting time of observation i t_{1i}	Ending time of observation i t_{2i}	Total demand (Units) D_i
1	0.0	1.4	308
2	4.2	5.0	220
3	8.8	9.4	187
4	11.8	16.0	1530
5	18.0	19.6	725
6	22.4	24.0	783

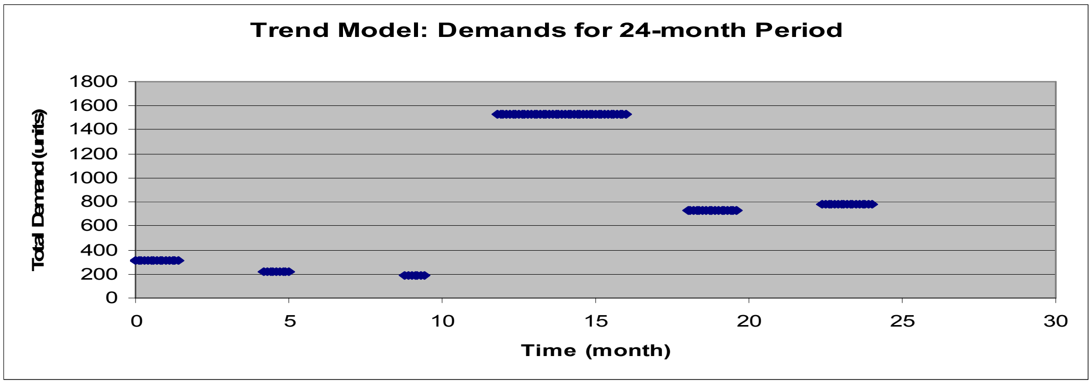

Figure 1. Trend Model: historical demands for 24-month period when $i = 1$ to 6

Step II

Obtain $\hat{D}_i$ for all i by using Equation (2.2). An illustration how to calculate $\hat{D}_i$, when $i = 1$ to 6, is shown in Table 4.

Table 4. Trend Model: calculation of estimate of total actual demand $\hat{D}_i$, when $i = 1$ to 6

Obser-vation	Starting time of obser-vation i	Ending time of obser-vation i	Total demand (Units)	Estimate of total demand
i	t_{1i}	t_{2i}	D_i	$\hat{D}_i = \hat{a}(t_{2i} - t_{1i}) + \frac{\hat{b}}{2}(t_{2i}{}^2 - t_{1i}{}^2)$
1	0.0	1.4	308	$\hat{D}_1 = \hat{a}(1.4 - 0.0) + \frac{\hat{b}}{2}(1.4^2 - 0.0^2) = 1.4\,\hat{a} + 0.98\,\hat{b}$
2	4.2	5.0	220	$\hat{D}_2 = \hat{a}(5.0 - 4.2) + \frac{\hat{b}}{2}(5.0^2 - 4.2^2) = 0.8\,\hat{a} + 3.68\,\hat{b}$
3	8.8	9.4	187	$\hat{D}_3 = \hat{a}(9.4 - 8.8) + \frac{\hat{b}}{2}(9.4^2 - 8.8^2) = 0.6\,\hat{a} + 5.46\,\hat{b}$
4	11.8	16.0	1530	$\hat{D}_4 = \hat{a}(16.0 - 11.8) + \frac{\hat{b}}{2}(16.0^2 - 11.8^2) = 4.2\,\hat{a} + 58.38\,\hat{b}$
5	18.0	19.6	725	$\hat{D}_5 = \hat{a}(19.6 - 18.0) + \frac{\hat{b}}{2}(19.6^2 - 18.0^2) = 1.6\,\hat{a} + 30.08\,\hat{b}$
6	22.4	24.0	783	$\hat{D}_6 = \hat{a}(24.0 - 22.4) + \frac{\hat{b}}{2}(24.0^2 - 22.4^2) = 1.6\,\hat{a} + 37.12\,\hat{b}$

Step III

Based on information in Table 4, write equation in matrix form to solve for $\hat{a}$ and $\hat{b}$.

$$\hat{D}_i = \hat{a}(t_{2i} - t_{1i}) + \frac{\hat{b}}{2}(t_{2i}^{\ 2} - t_{1i}^{\ 2})$$

$$\begin{bmatrix} 308 \\ 220 \\ 187 \\ 1530 \\ 725 \\ 783 \end{bmatrix} = \begin{bmatrix} 1.4 & 0.98 \\ 0.8 & 3.68 \\ 0.6 & 5.46 \\ 4.2 & 58.38 \\ 1.6 & 30.08 \\ 1.6 & 37.12 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}$$

Using Equation (2.9), therefore,

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 25.72 & 360.308 \\ 360.308 & 5735.24 \end{bmatrix}^{-1} \begin{bmatrix} 1.4 & 0.8 & 0.6 & 4.2 & 1.6 & 1.6 \\ 0.98 & 3.68 & 5.46 & 58.38 & 30.08 & 37.12 \end{bmatrix} \begin{bmatrix} 308 \\ 220 \\ 187 \\ 1530 \\ 725 \\ 783 \end{bmatrix}$$

Finally, we have

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 199.96 \\ 12.25 \end{bmatrix}$$

If someone wants to forecast the demand for January 2007, he/she can follow this calculation:

$$\hat{D}_7 = \hat{a} + \hat{b}t$$

Note: t=25 for January 2007, and therefore,

$$\hat{D}_7 = 199.96 + (12.25 \times 25)$$

$$\hat{D}_7 = 506.2 \text{ units}$$

3. COMPUTER SIMULATION AND RESULTS

This section demonstrates how a computer program is created to test the Trend Model. By using simulation technique, the data are generated in different situations associated with the change in number of observations and the percentage of missing data. The models are then employed to forecast on the data. Finally, the forecast errors and the COV are evaluated. The purpose is to explore how the number of observations and the percentage of missing data effect the performance of forecasting. The next section discusses the details of each procedure shown in Figure 2 and how the computer program is written.

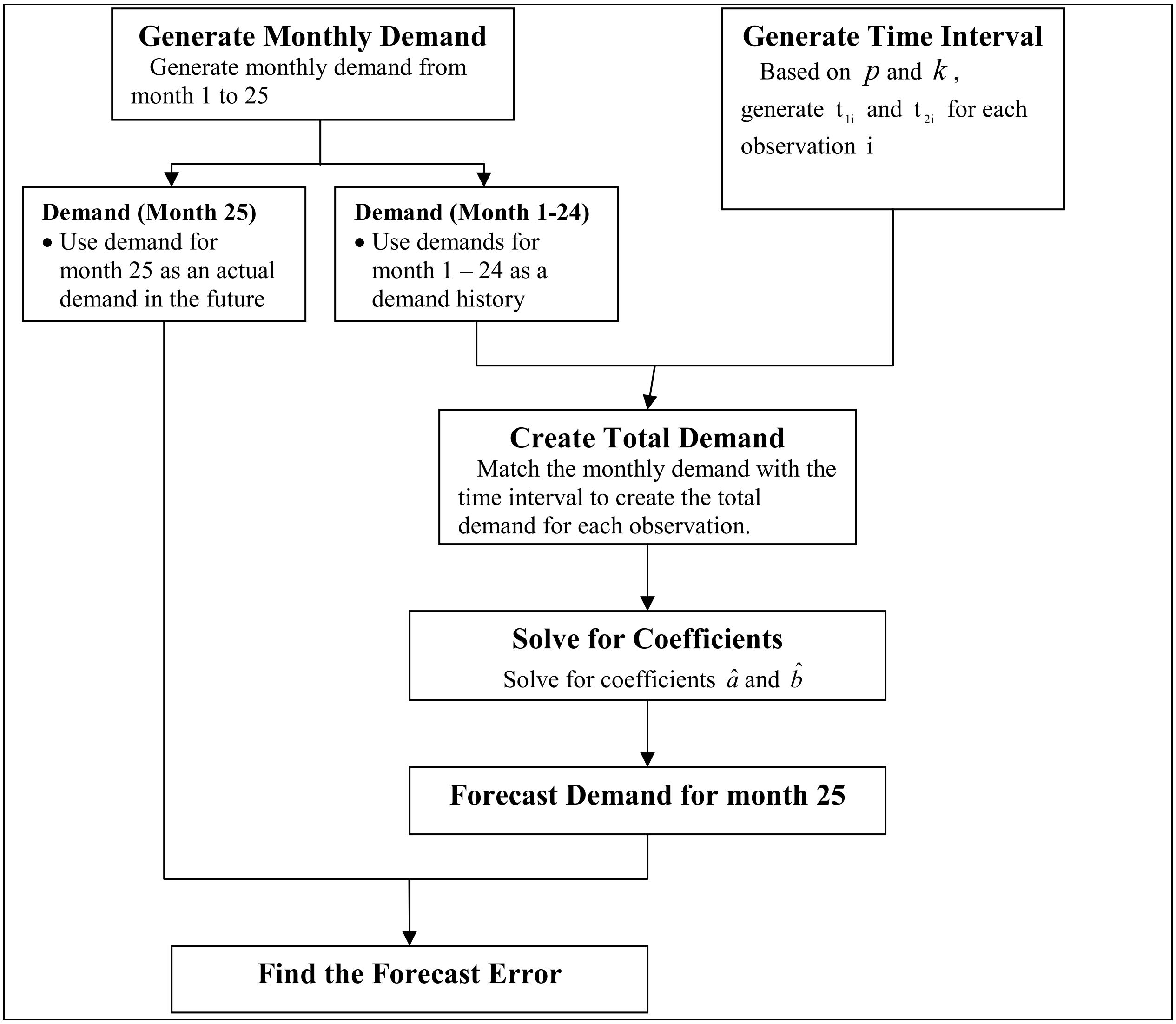

Figure2. Process of generating demands, forecasting, and finding the forecast error.

3.1 Data Generation Process

In order to test the model, the monthly demands for 25 months are generated. The demands for the first 24 months are used as a set of historical demands while the demand of month 25 is used as an actual demand in the future. This demand of month 25 will later be compared to the demand forecast of month 25 based on demand history from the first 24 months. To generate demand that seems to be realistic, parameters are set as explained in Figure 3.

<u>Trend Model</u> $\mu_t = a + bt$, let $a = 100$, $b = 10$ and $COV = 0.3$. The standard deviation, σ, can be calculated by $\sigma = COV \times \mu_t$. Then the computer program can be set that for $t = 1$ to 25, $D_t = \mu_t + Z\sigma$ where $Z = N(0,1)$.

Figure3. Illustration of data generation process to get 25-month demands for Trend Model

To find out how the number of observations and the percentage of missing data have an effect on the forecasting result, two parameters are set. First, set the number of observations, *k*. Four values of *k* have been selected as 6, 12, 18 and 24. Next, set the percentage of missing data, *p*, where five values of *p* are chosen as 0, 20, 40, 60, and 80 percent. In the program, the numbers of 0, 0.2, 0.4, 0.6, and 0.8 are used. For example, $p = 0.2$ means that 20 percent of time horizon is missing, which means 80 percent of the time horizon is available. In other words, if the total time horizon is 24 months, missing 20 percent means the time horizon available is (24)(1.0-0.2) = 19.2 months.

The next step is to find the starting time, t_{1i}, and the ending time, t_{2i}, of each observation, i, based on the number of observations and the percentage of missing data. For example, if the number of observation is six and 20 percent of the data is missing, then $k = 6$ and $p = 0.2$. We need to find t_{1i} and t_{2i} from $i = 1$ to 6 with minimum of $t_{11} = 0$, maximum of $t_{2k} = 24$, and $t_{2i} > t_{1i}$. The value of t_{1i} and t_{2i} must also satisfy the condition that $\sum_{i=1}^{k}(t_{2i} - t_{1i}) = (1 - p)(24)$.

Since $p = 0.2$, the calculation is illustrated as follows;

$$\sum_{i=1}^{6}(t_{2i} - t_{1i}) = (1\text{- }0.2)(24) = 19.2 \text{ months}$$

Since we have historical monthly demand for 24 months and the time interval for each observation, we can match the monthly and the time interval to create the total demand for each observation. The formulation is based on equation (3.1).

$$D_i = \sum_{\tau=\lceil t_{1i}\rceil}^{\lceil t_{2i}\rceil} d_\tau - d_{\lceil t_{1i}\rceil}(t_{1i} - \lfloor t_{1i}\rfloor) - d_{\lceil t_{2i}\rceil}(\lceil t_{2i}\rceil - t_{2i}) \qquad (3.1)$$

Where

D_i = total demand of observation i

d_τ = monthly demand of month τ which starts from time $\tau - 1$ to τ

t_{1i} = the starting point at the time of observation i

t_{2i} = the ending point at the time of observation i

$\lceil t_{1i}\rceil$ = round up operation of t_{1i}

$\lfloor t_{1i}\rfloor$ = round down operation of t_{1i}

For example, the first observation has the time interval from time 0.5 to 3.6. Monthly demands from month one to month four are 100, 195, 205, and 250 units respectively as shown in Figure 3.

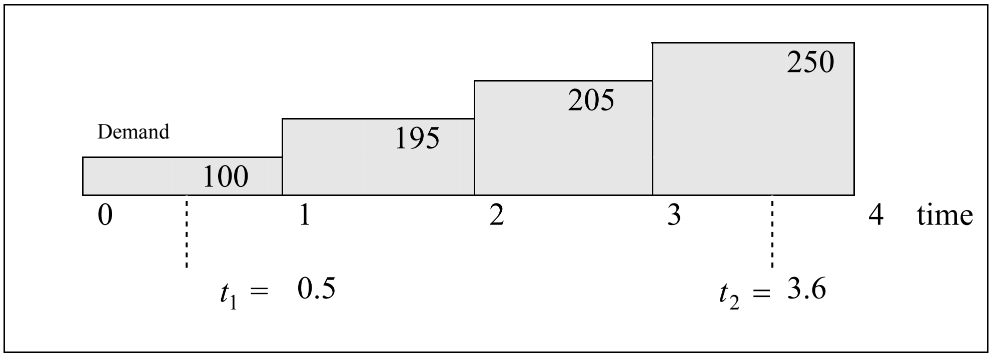

Figure 4. Matching the monthly demand with the time interval to create the total demand for each observation.

In order to get the total demand for this observation by using Equation (3.1), the calculation is shown as follows;

Since $t_{1i} = 0.5$ and $t_{2i} = 3.6$,

$$D_i = \sum_{\tau=\lceil 0.5\rceil}^{\lceil 3.6\rceil} d_\tau - d_{\lceil 0.5\rceil}(0.5 - \lfloor 0.5\rfloor) - d_{\lceil 3.7\rceil}(\lceil 3.6\rceil - 3.6)$$

$$= \sum_{\tau=1}^{4} d_\tau - d_1(0.5-0) - d_4(4-3.6)$$

$$= (100 + 195 + 205 + 250) - 100(0.5) - 250(0.4)$$

$$D_i = 600 \text{ units}$$

Therefore, the total demand for this observation is 600 units. We go on to the next observation and calculate the total demand in the same method until we have the total demand for each observation.

3.2 Solving for Coefficients and Forecasting

After all data are generated, the next step is to solve for coefficients, $\hat{a}$ and $\hat{b}$, needed for Trend Model. The coefficients $\hat{a}$ and $\hat{b}$ are then used to forecast the demand for month 25. The illustration of how to solve for the coefficients and use them to forecast can be found in Example1.

3.3 Measuring Forecasting Performance

In order to measure the performance of the forecasting for month 25, the forecasting error, e, needs to be found. Up to this point, the process of generating demands, forecasting, and finding the forecast error only shows one case of simulation. In this study, 1000 cases of simulation are generated for each combination of the number of observations and the percentage of missing data. This forecasting model has four types of *k,* the number of Observations (6, 12, 18, and 24) and five types of *p,* the percentage of missing data (0, 20, 40, 60, and 80 percent). Therefore, there are $4 \times 5 \times 1000 = 20{,}000$ cases of simulation. Table 5 shows the number of cases needed with the change in the number of observations and the percentage of missing data.

Table 5. Number of cases required with the change in the number of observations and the percentage of missing data for Trend Model

		Number of Observations, *k*				
		6	**12**	**18**	**24**	**Total**
Percentage of missing data, *p*	**0.8**	1000	1000	1000	1000	4000
	0.6	1000	1000	1000	1000	4000
	0.4	1000	1000	1000	1000	4000
	0.2	1000	1000	1000	1000	4000
	0.0	1000	1000	1000	1000	4000
Total		5000	5000	5000	5000	20000

After calculating the forecast errors from 1000 cases of each combination of k and p, we can find the sum of the squared errors (*SSE)* and the standard deviation of forecast errors, σ_e. An example of the calculation for the forecast errors and the sum of the squared errors is shown in Table 6.

Table 6. Illustration of calculation for the forecasting errors and the sum of the squared errors with 1000 cases of simulation.

Simulation Number	Actual Demand of month 25 for simulation number *j*	Estimated Demand of month 25 for simulation number *j*	Forecast Error, *e*	Square of the error, e^2
j	$D_{t=25,\,j}$	$\hat{D}_{t=25,\,j}$	$D_{t=25,j} - \hat{D}_{t=25,j}$	$(D_{t=25,j} - \hat{D}_{t=25,j})^2$
1	$D_{t=25,\,1}$	$\hat{D}_{t=25,\,1}$	$D_{t=25,\,1} - \hat{D}_{t=25,\,1}$	$(D_{t=25,\,1} - \hat{D}_{t=25,\,1})^2$
2	$D_{t=25,\,2}$	$\hat{D}_{t=25,\,2}$	$D_{t=25,\,2} - \hat{D}_{t=25,\,2}$	$(D_{t=25,\,2} - \hat{D}_{t=25,\,2})^2$
3	$D_{t=25,\,3}$	$\hat{D}_{t=25,\,3}$	$D_{t=25,\,3} - \hat{D}_{t=25,\,3}$	$(D_{t=25,\,3} - \hat{D}_{t=25,\,3})^2$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
1000	$D_{t=25,\,1000}$	$\hat{D}_{t=25,1000}$	$D_{t=25,\,1000} - \hat{D}_{t=25,1000}$	$(D_{t=25,\,1000} - \hat{D}_{t=25,1000})^2$
	$\sum_{j=1}^{1000} D_{t=25,\,j}$			$\sum_{j=1}^{1000} (D_{t=25,\,j} - \hat{D}_{t=25,\,j})^2$

Then, the standard deviation of forecast error is calculated by Equation (3.2).

$$\sigma_e = \sqrt{\frac{\sum_{j=1}^{1000}(D_{t=25,\, j} - \hat{D}_{t=25,\, j})^2}{N}} \tag{3.2}$$

where $D_{t=25,\, j}$ = actual demand of month 25 for simulation number j

$\hat{D}_{t=25,\, j}$ = estimated demand of month 25 for simulation number j

$\sum_{j=1}^{1000}(D_{t=25,\, j} - \hat{D}_{t=25,\, j})^2$ is the sum of the squared errors and N is the number of cases, which is 1000. So, the value of the standard deviation of forecast errors can be calculated as follow;

$$\sigma_e = \sqrt{\frac{\sum_{j=1}^{1000}(D_{t=25,\, j} - \hat{D}_{t=25,\, j})^2}{1000}}$$

The final step is to find the coefficient of variation of forecast errors, *COV*, which is calculated by using Equation (3.3).

$$COV = \frac{\sigma_e}{\overline{D}} \tag{3.3}$$

where $\overline{D} = (\sum_{j=1}^{1000} D_{t=25,\, j})/1000$

and $\overline{D}$ = average of actual monthly demand of month 25

The value of COV indicates the performance of the forecasting and is the relative measurement of one-period-ahead forecasting. The lower the number is, the more accurate the forecasting performs. In the next section, all simulation results and the impact of the difference of the number of observations and the percentage of missing data will be discussed.

3.4 Simulation Results

The results of simulation are then summarized in Table 7. The results from Table 7 are then plotted in charts in Figure 5. Each value of COV is calculated from 1,000 cases for each combination of *p* and *k,* where *p* is the percentage of missing data and *k* is the number of observations.

Table 7. Trend Model: COV with percentage of missing data and number of observations

COV	Number of Observations *k*				
% missing *p*	6	12	18	24	Average of Row
0.8	0.3534	0.3365	0.3243	0.3181	0.3331
0.6	0.3327	0.3286	0.3224	0.3136	0.3244
0.4	0.3186	0.3163	0.3148	0.3148	0.3161
0.2	0.3129	0.3105	0.3097	0.3093	0.3106
0	0.3094	0.3081	0.3096	0.3069	0.3085
Average of Column	0.3254	0.3200	0.3162	0.3126	0.3185

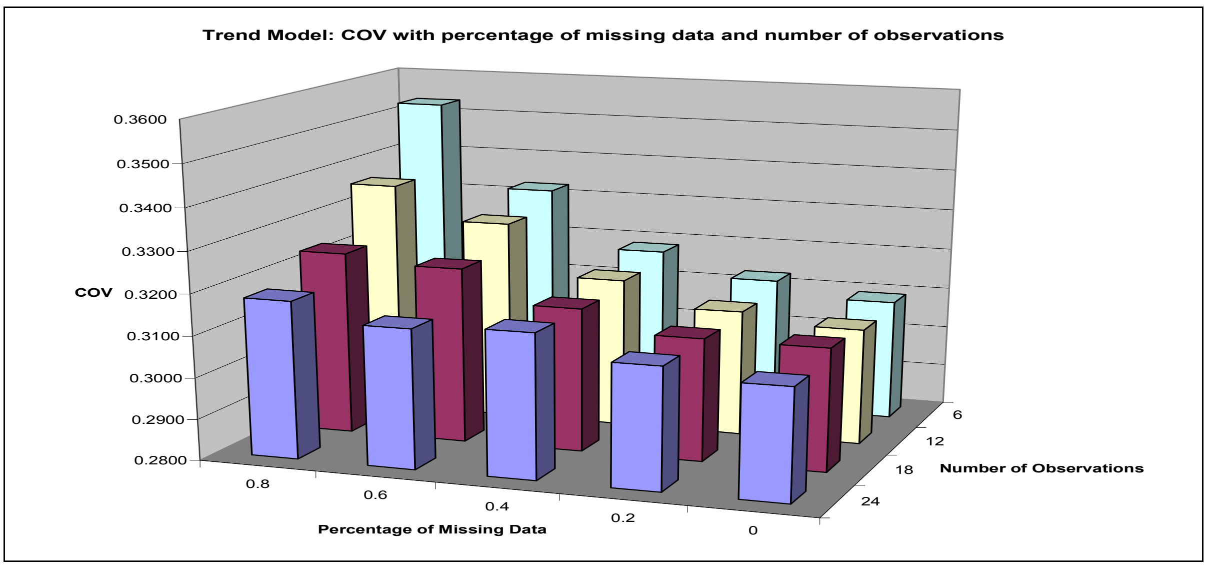

Figure 5. Trend Model: COV with percentage of missing data and number of observations

By inspection, we found that the outputs perform in the same pattern. The numbers in the right column in the Table 7 under the topic “Average of row” indicate that when the percentage of missing data decreases, the COV also decreases which means that the forecasting is more accurate. The numbers in the bottom row of Table 7 under the topic “Average of column” also indicate that when the number of observations increases, the value of COV decreases.

The lowest value of COV is 0.3069. If we divide each value of COV in Table 7 by 0.3069, we have the results as shown in Table 8.

Table 8. Trend Model: forecasting performance ratio, the value of COV for each combination of *p* and *k* divided by the lowest value of COV from Table 7

Ratio	**Number of Observations *k***			
% missing *p*	**6**	**12**	**18**	**24**
0.8	1.1513	1.0964	1.0567	1.0363
0.6	1.0840	1.0707	1.0505	1.0218
0.4	1.0380	1.0307	1.0255	1.0256
0.2	1.0196	1.0117	1.0089	1.0078
0	1.0080	1.0039	1.0088	1.0000

Table 8 shows that the forecasting performs best when the percentage of missing data is zero. When the number of observations is six and 80 percent of data is missing, the ratio is 15.13 percent worse than when the number of observations is 24 and the percentage of missing data is zero. We conclude that the higher the number of observations is, the more accurate the forecasting performs. The results also show that the forecasting performs better with the lower percentage of missing data.

4. CONCLUSION

In this study, a forecasting technique, in dealing with unequal time intervals and missing demand histories, is provided. The coefficient of variation of the forecast error (COV) is used to measure the performance of forecasting. The model copes with the trend demand pattern. The approach resembles the Linear Regression Model. The matrix approach and the method of ordinary least squares (OLS) are used to estimate the coefficients needed to forecast demand in the future time period. The results show that this forecasting technique for Trend Model performs better with a higher number of observations and a lower percentage of missing data.

References

Aldrin, M. and E. Damsleth (1989). Forecasting Non-Seasonal Time Series with Missing Observations. *Journal of Forecasting,* (8)2, 97-116.

Brown, R. G. (1963). *Forecasting and Prediction of Discrete Time Series*. New Jersey: Prentice-Hall.

Dunsimuir, W. and P. M. Robinson (1981). Estimation of Time Series Models in the Presence of Missing Data. *Journal of the American Statistical Association* (76)375, 560-568.

Gujarati, D. M. (1995). *Basic Econometrics*. (3rd ed.). McGraw-Hill.

Manaspiti, P. (1990). *Forecasting on Implicit Demands*. Doctoral Dissertation, Illinois Institute of Technology , Chicago, Illinois.

Ruangkanjanases, A. (2007) *Forecasting with Irregular Demand History*. Doctoral Dissertation, Illinois Institute of Technology, Chicago, Illinois.

Thomopoulos, N. T. (1980). *Applied Forecasting Methods*. Englewood Cliffs: Prentice-Hall.