EXTRAPOLATION-BASED GREY MODEL FOR SMALL-DATA-SET FORECASTING

Abstract. Product life cycles have become increasingly shorter owing to the rise of global competition in recent decades. Competitive tension is especially high in electronics-related industries. It is usually difficult for most enterprises to collect sufficient quantities of samples with which to obtain useful information when making decisions in such a highly competitive environment. Grey system theory plays a vital role in addressing the issue of insufficient sample quantities. The traditional GM(1,1) model is well known for its ability to generate useful forecasts with a small quantity of samples; however, the newest datum is always weakened to alleviate the randomness of data in the traditional GM(1,1) model, causing it to output higher prediction errors. To overcome such imperfections, this study proposes a modified grey forecasting model named EP-GM(1,1), in which a new equation for calculating the background values in the traditional GM(1,1) model is developed based on linear extrapolation to emphasize the importance of the newest datum. To evaluate the forecasting ability of EP-GM(1,1), the monthly demand of thin-film-transistor liquid-crystal display panels were employed for experimentation. The results indicate that EP-GM(1,1) can engender a favorable prediction result, demonstrating that the model is a feasible tool for small-sample forecasting.

Keywords: Grey system theory, Small data set; Forecasting; Short-term demand.

JEL Classification: C01, C02, C13, C22, C53, M11

DOI: 10.24818/18423264/53.1.19.11

* Corresponding author.
1. Introduction

Product life cycles are subject to increasing global competition, and thus, have become increasingly shorter in recent years (Li et al., 2010), forcing companies to contend with a rapidly changing industry environment (Chang et al., 2016). Therefore, producing various customized products is a current trend for manufacturing industries, compelling engineers and managers to shorten product development periods (Krajewski et al., 2010). However, one pending issue must be overcome, namely, the need to derive useful information from limited and statistically insufficient data within a short time period (Lin & Tsai, 2014). This issue engenders difficulties in ensuring the quality of mass production in theory and practice (Chang et al., 2015). Therefore, the question of how to appropriately employ small data sets to improve manufacturing performance is crucial for maintaining industrial competitive advantages (Li et al., 2012).

Statistical theory and machine-learning algorithms are widely adopted for extracting management knowledge, however, these approaches are theoretically developed based on large samples (Lind et al., 2013; Witten et al., 2011). When a sample size is small, these methods do not easily yield favorable predictions (Liu & Lin, 2006). Therefore, the extraction of valuable information with limited data for assisting engineers and managers in decision-making holds significant practical value.

Grey system theory is capable of characterizing an unknown system using small samples without the need for conditions in conformity with statistical assumptions (Deng, 1982), thereby providing a strong technical backing for small-data-set analysis (Pearman, 2016). Grey models require only a small number of samples to achieve reliable and acceptable forecasting accuracy (Liu & Lin, 2006), and have been successfully adopted in various fields with favorable outcomes (Boran, 2015; Camelia et al., 2013; Chang et al., 2013; Evans, 2014; Salmeron, 2016; Wang & Hao, 2016; Zeng et al., 2016).

The traditional GM(1,1) model is one of the most popular grey models for time series forecasting, as well as being a critical core of grey approaches. In recent years, several experts have modified the original GM(1,1) model to improve its forecasting accuracy (Wu et al., 2013; Xie & Liu, 2009; Yamaguchi et al., 2007). However, the model could still be further improved. Therefore, this study proposes a new calculation formula for the background value in grey modeling, which plays a crucial role in the traditional grey model, thereby constructing an extrapolation-based grey model called EP-GM(1,1). This new method can improve forecasting results based on small data sets.

To confirm the validity of the proposed model, this study employed one real data set to conduct an empirical analysis for evaluating the forecasting ability and practical value of the model, namely the total demand of thin-film-transistor liquid-crystal display (TFT LCD) panels provided by a leading manufacturer in Taiwan. The empirical results indicate that EP-GM(1,1) can yield favorable
Extrapolation-based Grey Model for Small-data-set Forecasting

outcomes, and thus, is a feasible small-data-set forecasting method for engineers and managers.

The remainder of this paper is organized as follows: Section 2 introduces the traditional GM(1,1) model, as well as the theoretical concept and modeling procedure of EP-GM(1,1). Section 3 examines the forecasting performance and describes a comparison of EP-GM(1,1) through the application of one real case. Finally, Section 4 discusses the outcomes and presents conclusions.

2. Methodology

When the sample size of the obtained data set is small, how to develop a feasible modeling procedure to predict possible future trends is a critical issue for researchers. Because of insufficient information, it is difficult to grasp such situation using statistical methods and data mining techniques. The grey model can perform modeling and forecasting procedures with only a small number of samples, and can also yield favorable prediction results (Liu & Lin, 2006). Therefore, this study employed the traditional GM(1,1) model as a basis for developing an improved grey model, EP-GM(1,1), for solving small data set forecasting problems.

2.1 Traditional GM(1,1) model

Of all the grey forecasting models, the first-order one-variable grey model, GM(1,1), is the most frequently employed. For this model, only a small number of samples are required to yield a valuable forecasting result. Its main functions are to employ data mapping to transform the state space, diminish the randomness of data, and identify hidden regular patterns to build a model for small-data-set forecasting (Deng, 1989). It is thus a crucial method for overcoming small-sample forecasting issues. The computational steps of the traditional GM(1,1) model are as follows:

**Step 1:** Consider the original and nonnegative data series \( X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\} \), \( n \geq 4 \), where \( x^{(0)}(k) \) represents the \( k \)th phase of the data.

**Step 2:** Perform the accumulated generating operation to identify the potential hidden regularity to generate a new data series, \( X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)\} \).

\[
x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(1), \quad k = 1, 2, \ldots, n \tag{1}
\]

**Step 3:** Compute the background values \( Z^{(1)} = \{z^{(1)}(2), z^{(1)}(3), \ldots, z^{(1)}(n)\} \) by using Equation (2).

\[
z^{(1)}(k) = \frac{1}{2} x^{(1)}(k-1) + x^{(1)}(k), \quad k = 2, 3, \ldots, n \tag{2}
\]

**Step 4:** Establish the grey differential equation.
\[ x^{(0)}(k) + ax^{(0)}(k) = b \]  

**Step 5:** Extend Equation (3) as the vector-matrix form.

\[
\begin{bmatrix}
  x^{(0)}(2) \\
  x^{(0)}(3) \\
  \vdots \\
  x^{(0)}(n)
\end{bmatrix} = \begin{bmatrix}
  -z^{(i)}(2) \\
  -z^{(i)}(3) \\
  \vdots \\
  -z^{(i)}(n)
\end{bmatrix} \begin{bmatrix}
  a \\
  b
\end{bmatrix}
\]  

(4)

Let

\[ Y = \begin{bmatrix} x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n) \end{bmatrix}^T, \quad \hat{a} = [a, b]^T, \quad \text{and} \quad B = \begin{bmatrix}
  -z^{(i)}(2) & 1 \\
  -z^{(i)}(3) & 1 \\
  \vdots & \vdots \\
  -z^{(i)}(n) & 1
\end{bmatrix} \]  

(5)

**Step 6:** Determine the developing coefficient \( a \) and grey input \( b \) by using the ordinary least square method.

\[ \hat{a} = \left(B^T B\right)^{-1} B^T Y \]  

(6)

**Step 7:** Solve the ordinary first-order differential equation \( dx^{(i)}(t)/dt + ax^{(i)}(t) = b \) with the initial condition \( x^{(0)}(1) = x^{(0)}(1) \) to establish the grey forecasting model.

\[
\begin{align*}
  \hat{x}^{(i)}(k + 1) &= \left(x^{(0)}(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a} \\
  \hat{x}^{(i)}(k + 1) &= \hat{x}^{(i)}(k + 1) - \hat{x}^{(i)}(k)
\end{align*}
\]  

(7)

**Step 8:** Obtain the desired forecasting output through the described model.

2.2 New calculation formula of the background value

Previous studies have revealed that the background value is crucial in grey modeling, and usually directly influences the applicability and forecasting performance of the model. Various background values influence the developing coefficient and grey input to be determined, and exert a direct impact on the forecasting accuracy of the model. The basic function of the background value is to alleviate the fluctuation of the time series by smoothing the data, as well as explore the components that cause the regular changes in time series. When Equation (2) is rewritten as Equation (8), the newest datum is evidently weakened to alleviate the randomness of the data. Although this method is easy and effective, it is vital that the newest datum not be considered concurrently because this may result in a more substantial prediction error. According to the fifth axiom of grey systems: the principle of new information priority, the function of new pieces of information is greater than those of old pieces of information (Liu & Lin, 2006. Therefore,
Extrapolation-based Grey Model for Small-data-set Forecasting

developing a new calculation formula for the background values that is able to smooth data to alleviate randomness and emphasize the importance of the newest datum is worthwhile.

\[ z^{(1)}(k) = x^{(1)}(k) - \frac{1}{2} x^{(0)}(k), \quad k = 2, 3, \ldots, n \] (8)

The background value \( z^{(1)}(k) \) is employed to replace the accumulated value, \( x^{(1)}(k) \); therefore, its values must surround \( x^{(1)}(k) \). The traditional method is to apply the mean of \( x^{(1)}(k-1) \) and \( x^{(1)}(k) \) as the background value \( z^{(1)}(k) \). The background value \( z^{(1)}(k) \) in the traditional GM(1,1) model is one of the linearly interpolated values between \( x^{(1)}(k-1) \) and \( x^{(1)}(k) \). Linear extrapolation is an alternative numerical approximation method for alleviating the shortcoming inherent in the weakened newest datum.

The extended value of \( x^{(1)}(k) \) obtained through linear extrapolation is still approximate to the original value, able to strengthen the newest datum, and easy to calculate, thereby satisfying the basic conditions for selecting a background value. This study thus adopts a new calculation formula for the background value based on linear extrapolation. Its relative position is presented in Figure 1. If the extended coefficient is set as alpha, the proposed new \( z^{(1)}(k) \) could be obtained with Equation (9) through the ratio rule. Here, for easy application, the extended coefficient is set at a fix value (0.5), and subsequently, Equation (9) is simplified to Equation (10). From Equation (10), it is evident that discovering the new calculation formula for the background value can strengthen the newest datum rather than weaken it.

\[
\begin{align*}
\text{Traditional} & \quad \text{New} \\
\begin{array}{c}
\longrightarrow \quad x^{(1)}(k) \\
\longrightarrow \quad x^{(1)}(k) \\
\longrightarrow \quad x^{(1)}(k-1) \\
\longrightarrow \quad 0.5 \\
\longrightarrow \quad \alpha \\
\longrightarrow \quad k-1 \\
\longrightarrow \quad k \\
\longrightarrow \quad k+\alpha \\
k \\
\end{array}
\end{align*}
\]

Figure 1: New calculation formula of the background value.

175
\[ z^{(l)}(k) = (1 + \alpha)\left[ x^{(l)}(k) - x^{(l)}(k-1)\right] + x^{(l)}(k-1), \ k = 2,3,\ldots,n \]  
\[ z^{(l)}(k) = x^{(l)}(k) + \frac{1}{2} x^{(0)}(k), \ k = 2,3,\ldots,n \]  

2.3 Modeling procedure of EP-GM(1,1)

The background value of the grey model is a crucial factor that affects forecasting performance. To emphasize the importance of the newest datum, we propose a new calculation formula for the background value based on the linear extrapolation. The modified grey model derived from this calculation formula is called EP-GM(1,1). The modeling procedure of EP-GM(1,1) is as follows:

**Step 1:** Consider the original and nonnegative data series \( X^{(0)} = \{ x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n) \} \).

**Step 2:** Produce the accumulated generating series \( X^{(l)} = \{ x^{(l)}(1), x^{(l)}(2), \ldots, x^{(l)}(n) \} \) through Equation (1).

**Step 3:** Compute the background values \( Z^{(l)} = \{ z^{(l)}(2), z^{(l)}(3), \ldots, z^{(l)}(n) \} \) by using Equation (10).

**Step 4:** Establish the grey differential equation, and determine the developing coefficient and grey input using Equations (3)-(6).

**Step 5:** Create the EP-GM(1,1) model and use it to obtain the desired forecasting outputs.

3. Empirical results

Empirical analysis was conducted using one real data set to compare the forecasting ability of EP-GM(1,1) with that of the traditional GM(1,1) model and the back propagation neural network (BPNN). Detailed descriptions are provided in the following subsections.

3.1 Total demand data of TFT LCD panels

The TFT LCD industry is a mature industry characterized by intense competition; its supply is greater than its demand because of China’s active participation in production. Under such conditions, manufacturers must reduce inventory levels and improve inventory turnover rates.

TFT LCD panel manufacturing is a highly capital-intensive industry, and the production quantity of TFT LCD panels can heavily influence production costs. Economies of scale are the primary competitive strategy adopted by manufacturers in the TFT LCD industry. They increase production quantities to reduce production...
unit costs. However, an increase in production quantity cannot always reduce manufacturing costs; when the production quantity exceeds a critical value, manufacturing unit costs (e.g., management costs, warehousing costs, taxes, interest and insurance) may increase. Therefore, the production quantity must be maintained at a suitable balance point in consideration of the total cost, and an accurate short-term demand forecast is necessary for meeting the requirements of coordinating production and marketing.

Because the global business cycle is a crucial factor affecting the demand for TFT LCD panels, and the change of business cycles has significantly increased in recent years, a prediction based on vast amounts of long-term historical data does not satisfy the requirements of the TFT LCD industry. Therefore, short-term demand must be forecast accurately to overcome the challenges of coordinating production and marketing.

This study adopted data provided by a leading TFT LCD panel manufacturer in Taiwan to confirm the forecasting ability and practical value of the EP-GM(1,1) model. The data set is the manufacturer’s total demand for TFT LCD panels, comprising 36-period time series data ranging from January 2010 to December 2012. Because of confidentiality requirements, all data were transformed to fall within the $[1, 2]$ interval using min-max normalization in advance. The normalized data are presented in Table 1. Equation (11) is the formula adopted in this study to carry out min-max normalization, in which $x_{\text{max}}$ is the maximum value in the pre-normalized data set, $x_{\text{min}}$ is the minimum value in the pre-normalized data set, $x^{(0)}(k)$ is the pre-normalized datum, and $v^{(0)}(k)$ is the post-normalized datum. In this study, every time four data were employed for model fitting and the next datum $\hat{x}(5)$ was employed for ex post testing. A total of 32 models and 32 forecasting outputs were obtained.

$$v^{(0)}(k) = \frac{x^{(0)}(k) - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} + 1$$

### Table 1. Monthly demand of TFT LCD panels

<table>
<thead>
<tr>
<th>Months</th>
<th>Demands</th>
<th>Months</th>
<th>Demands</th>
<th>Months</th>
<th>Demands</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010/01</td>
<td>1.135</td>
<td>2010/02</td>
<td>1.000</td>
<td>2010/03</td>
<td>1.231</td>
</tr>
<tr>
<td>2010/03</td>
<td>1.277</td>
<td>2010/04</td>
<td>1.274</td>
<td>2010/05</td>
<td>1.176</td>
</tr>
<tr>
<td>2010/04</td>
<td>1.587</td>
<td>2010/05</td>
<td>1.334</td>
<td>2010/06</td>
<td>1.850</td>
</tr>
<tr>
<td>2010/05</td>
<td>1.835</td>
<td>2011/01</td>
<td>1.851</td>
<td>2011/02</td>
<td>1.851</td>
</tr>
<tr>
<td>2010/06</td>
<td>1.442</td>
<td>2011/02</td>
<td>1.696</td>
<td>2011/03</td>
<td>1.718</td>
</tr>
<tr>
<td>2011/01</td>
<td>1.446</td>
<td>2011/03</td>
<td>1.718</td>
<td>2011/04</td>
<td>1.728</td>
</tr>
<tr>
<td>2011/02</td>
<td>1.696</td>
<td>2012/01</td>
<td>1.718</td>
<td>2012/02</td>
<td>1.714</td>
</tr>
<tr>
<td>2011/03</td>
<td>1.718</td>
<td>2012/03</td>
<td>1.728</td>
<td>2012/04</td>
<td>1.714</td>
</tr>
</tbody>
</table>
3.2 EP-GM(1,1) modeling example

The first four monthly demand value in the TFT LCD case were selected as examples for creating a forecasting model for illustrating the calculation process of EP-GM(1,1). Specifically, the actual demand of panels from January 2010 to April 2010 was analyzed to establish a model for predicting panel demand in May 2010, (i.e., \{1.135,1.000,1.231,1.277\}). Through calculation, the modeling parameters \(a\) and \(b\) were determined to be \(-0.10507\) and \(0.75265\), respectively. Therefore, the forecasting model is \(\hat{x}(k+1) = 8.2983e^{0.10507k} - 7.1633\), and the next observation was predicted as \(\hat{x}(5) = 1.26\). The detailed computation procedure is presented in Table 2.

<table>
<thead>
<tr>
<th>Order</th>
<th>(\chi^{(0)}(k))</th>
<th>(\chi^{(1)}(k))</th>
<th>(\varepsilon^{(1)}(k))</th>
<th>(\hat{x}(k))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.135</td>
<td>1.135</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>1.000</td>
<td>2.135</td>
<td>2.6350</td>
<td>0.9193</td>
</tr>
<tr>
<td>3</td>
<td>1.231</td>
<td>3.366</td>
<td>3.9815</td>
<td>1.0212</td>
</tr>
<tr>
<td>4</td>
<td>1.277</td>
<td>4.643</td>
<td>5.2815</td>
<td>1.1343</td>
</tr>
<tr>
<td>5</td>
<td>1.274</td>
<td>-</td>
<td>-</td>
<td>1.2600</td>
</tr>
</tbody>
</table>

3.3 Forecasting accuracy comparison

Two popular approaches were selected for comparison with EP-GM(1,1) for evaluating forecasting ability. These were the traditional GM(1,1) model and the BPNN. The traditional GM(1,1) model is the most typical forecasting model in grey system theory because of its simplicity and commonality of use. The BPNN is a famous type of neural network that is also popular for its convenience in use. These two approaches also employed four observations to forecast the subsequent output.

Liu and Lin (2006) asserted that the forecasting error is a critical index for measuring the forecasting ability of a grey model. Yokum and Armstrong (1995) also indicated that accuracy is an essential criterion for testing the performance of a
Extrapolation-based Grey Model for Small-data-set Forecasting

forecasting approach. In the experiment, the mean absolute percentage error (MAPE) was employed to evaluate the performance of forecasting methods, because it can serve as a benchmark and is a stable method of error-based measurement (Makridakis, 1993). The MAPE criteria for evaluating a forecasting model are presented in Table 3 (DeLurgio, 1998). The MAPE is calculated as follows:

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^{m} \left| \frac{\hat{x}_i - x_i}{x_i} \right| \times 100\%$$

(12)

were $m$ is the sum of the testing samples, and $\hat{x}_i$ and $x_i$ are the estimated and actual values of the $i$th testing sample, respectively.

Table 3. MAPE criteria

<table>
<thead>
<tr>
<th>MAPE</th>
<th>Forecasting power</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10%</td>
<td>Highly accurate forecasting</td>
</tr>
<tr>
<td>10-20%</td>
<td>Good forecasting</td>
</tr>
<tr>
<td>20-50%</td>
<td>Reasonable forecasting</td>
</tr>
<tr>
<td>&gt;50%</td>
<td>Inaccurate forecasting</td>
</tr>
</tbody>
</table>

Table 4 presents comparisons of the three forecasting approaches. The empirical results indicate that the EP-GM(1,1) model obtains smaller errors than the other two approaches do, and its MAPE was 4.46%; falling within an acceptable range (Table 3). This indicates that EP-GM(1,1) exhibited a superior forecasting ability compared with the traditional GM(1,1) model and the BPNN. Furthermore, improvements in the MAPE of the proposed model in comparison with the traditional GM(1,1) model reached 59.38%, indicating that EP-GM(1,1) can alleviate the shortcomings of the traditional model to obtain superior forecasting results through consideration of the importance of the new datum.

Table 4. Performance among various grey models

<table>
<thead>
<tr>
<th>Methods</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP-GM(1,1)</td>
<td>4.46</td>
</tr>
<tr>
<td>GM(1,1)</td>
<td>10.98</td>
</tr>
<tr>
<td>BPNN</td>
<td>10.57</td>
</tr>
</tbody>
</table>

4. Conclusions and discussion

In the rapidly changing industry environment, manufacturers must continue to improve manufacturing performance. How to acquire useful information from limited and statistically insufficient data within a short time period is an unresolved issue for enterprises. To overcome the uncertainty of decision-making, a robust forecasting approach is indispensable to engineers and
managers. Good forecasting ability can not only lower costs but also enable an organization to be more efficient in operational management, thereby enhancing the company's competitiveness. However, in the face of global competition, collecting sufficient observations to understand future trends is not easy. For this reason, appropriately forecasting product demand by employing small samples is vital for enterprises to maintain industrial competitive advantages.

Grey system theory can build models on the basis of small data sets, thereby satisfying the requirements of enterprises. The traditional GM(1,1) model is the most popular grey model because of its ease of use. However, the newest datum is weakened to alleviate the randomness of data in the traditional GM(1,1), and this may result in a larger prediction error. Therefore, linear extrapolation was adopted in this study to emphasize the importance of the newest datum, and a new equation for the background value in grey modeling was developed to create the extrapolation-based grey model, EP-GM(1,1), for solving small-sample forecasting problems. Real case of the monthly demand of TFT LCD panels was employed to evaluate the forecasting performance of the EP-GM(1,1) model. The outcomes demonstrated that the proposed grey model performs well with small data sets. The EP-GM(1,1) model evidently possesses considerable practical value, as well as being a useful method for implementing short-term demand forecasts using limited samples.

In the future, the EP-GM(1,1) model could be applied to other fields, such as finance, transportation, and engineering to further confirm its effectiveness. A recommendation for further study is selection of the coefficient in the proposed model as a fixed value for convenience of use to obtain the optimum extended coefficient. Finally, methods of extending the modeling concept of EP-GM(1,1) to other types of grey model are also worthy of further study.

REFERENCES

Extrapolation-based Grey Model for Small-data-set Forecasting


182

Che-Jung Chang, Der-Chiang Li, Chien-Chih Chen, Wen-Chih Chen