Forecasting of financial series for the Nevada Department of Transportation using deterministic and stochastic methodologies

Hanns de la Fuente-Mella\textsuperscript{a,}\textsuperscript{*}, Alexander Paz-Cruz\textsuperscript{b}, Rebecca Conover\textsuperscript{b}, Alauddin Khan\textsuperscript{b}

\textsuperscript{a}Pontificia Universidad Católica de Valparaíso, Facultad de Ciencias Económicas y Administrativas, Escuela de Comercio, Avenida Brasil, 2830, Valparaíso, Chile
\textsuperscript{b}University of Nevada, Las Vegas, Department of Civil and Environmental Engineering, 4505 S Maryland Parkway, PO Box 454015, NV 89154, Las Vegas, United States

Abstract

In this research, a set of financial series was forecasted using data from the Nevada Department of Transportation (NDOT) with the objective of facilitating financial management and associated decision-making. Both deterministic and stochastic methods for seven financial time series, which were independent and univariate, from NDOT’s financial data warehouse. Data from 2001 to 2014 were annually and equally spaced. The data series included drivers’ license fees, federal aid revenue, gas tax revenue, motor carrier fees, registration fees, special fuel tax revenue, and total state revenue. The deterministic forecasting methods used included Simple, Holt, Brown, and Damped Trend, and the stochastic forecast used ARIMA (p, d, q). The Simple and Holt methods provided an adequate forecast for 28% of the cases. Brown’s method provided an adequate forecast for 44% of the cases. However, the stochastic process, the ARIMA method, did not find an acceptable goodness of fit. An absence of large datasets likely precluded an appropriate estimation when using the ARIMA (p, d, q) method. Tests were performed using functional curves including linear, logarithmic, inverse, quadratic, cubic, growth-exponential, and logistic. The best fits were obtained using the cubic functional form, with an average coefficient of determination of 82%.

Keywords: Transportation; Budget forecast; Stochastic methods; Deterministic methods; Functional forms fit

\textsuperscript{*}Corresponding author.
E-mail address: hanns.delafuente@ucv.cl
1. Introduction

Accurate revenue forecasting is becoming an even more significant aspect than ever for the financial management for public-sector agencies. In the absence of reliable revenue forecasts, financial management becomes problematic when true revenue is collected. Having an accurate financial forecast that utilizes sufficient historical data is required to reduce uncertainty and fiscal stress [1]. In addition, reliable and accurate revenue forecasting models facilitate making better decisions regarding fund allocation, project scheduling, and emergency management.

This study, which proposes various forecasting models, uses data from the Nevada Department of Transportation (NDOT). Potentially, other Departments of Transportation or public sector entities can use these methods. However, it should be noted that a method that best fits the sample data used in this study is not guaranteed to be the method that best fits other sets of time-series data. Moreover, the proposed methodology requires an analyst to determine the best model for the data by interpreting goodness of fit.

2. Methodology

Along with the models utilized in this research, hundreds of other potential modeling techniques exist [2]. Potential alternatives, for example, are hybrid models [2] and the method of averaging estimates from several different models [3]. The methods used in this research are among those commonly utilized for financial forecasting.

This study used seven financial time series that were independent and univariate. The data were annually and equally spaced between 2001 and 2014. The series included drivers’ licenses fees, federal aid, gas tax revenues, motor carrier fees, registration fees, special fuel-tax revenues, and total state revenues. Federal aid income included all monies given to the state by the federal government through the Federal Highway Administration (FHWA) with the purpose of funding qualified highway projects.

Gas taxes and special fuel taxes are the excise taxes levied on gasoline and other fuels—such as diesel, compressed natural gas, or propane—sold within the state. All drivers and vehicles within the State of Nevada are required to be licensed and registered through the Nevada Department of Motor Vehicles (DMV); thus, they incur drivers’ license fees and vehicle registration fees. Motor carrier fees, such as vehicle registration fees, are levied on large or commercial vehicles operating within the state.

Initially, an exploratory analysis was developed to determine the characteristics of each series. To predict each of the series, methods were used for deterministic exponential smoothing [4, 5], stochastic models, and autoregressive integrated moving average (ARIMA) models [6]. Each forecast result was compared using goodness of fit indices of the root-mean-square error (RMSE) and the mean absolute percentage error (MAPE). In the last step, different functional forms were used to adjust each one of the series, and the choice of best goodness of fit was based on the R-squared index.

2.1. Exploratory analysis

Exploratory empirical analysis was conducted on each of the seven series. Each case had the same number of observations, n = 14 (2001-2014), and descriptive analysis was performed over time for each of the seven series. As illustrated in Table 1, the largest amount of revenue collected by NDOT came from federal aid collected in 2012. The smallest amount of revenue was collected through drivers’ license fees in 2001.

2.2. Deterministic forecasting

The deterministic forecast for these seven series included 1) a Simple deterministic model, 2) the Holt model, 3) the Brown model, and the Damped Trend model, all of which are discussed in greater detail in this section. These models were used to calculate the results after six future periods (2015-2020).
A single moving average weights the observations equally. The methods used in this study for forecasting provided a smoothed time series [7]. Exponential smoothing assigned exponentially decreasing weights as the observations got older. In other words, recent observations received relatively more weight than older observations.

Exponential smoothing was first suggested by Brown [4] and expanded by Holt [5]. Eq. 1 is commonly used and attributed to Brown, and therefore known as Brown’s Simple Exponential Smoothing [4].

\[ A_t = \alpha X_t + (1 - \alpha)A_{t-1} \] (1)

where:
\( A_t \) = the attenuated average time series after observing \( X_t \)
\( X_t \) = observed values
\( \alpha \) = an attenuation constant

Simple exponential smoothing does not provide adequate estimates when there is a trend in the data. To address these cases, a number of methods were created that were known as double exponential smoothing, as shown in Eqs. 2 and 3. The basic idea behind double exponential smoothing is to introduce a term that takes into account the possibility of a series having a trend. This term is updated by means of exponential smoothing [7].

\[ A_t = \alpha X_t + (1 - \alpha)(A_{t-1} + T_{t-1}) \]
\[ T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1} \] (2)

where:
\( A_t \) = the attenuated average time series after observing \( X_t \)
\( T_t \) = the trend estimate
\( X_t \) = observed values
\( \alpha, \beta \) = attenuation constants

\[ A_t = \alpha X_t + (1 - \alpha)(A_{t-1} \ast T_{t-1}) \]
\[ T_t = \beta(A_t / A_{t-1}) + (1 - \beta)T_{t-1} \] (3)

where:
\( A_t \) = the attenuated average time series after observing \( X_t \)
\( T_t \) = the trend estimate
\( X_t \) = observed values
\( \alpha, \beta \) = attenuation constants
Forecasts provided by double exponential smoothing methods show a constant trend in the future. These methods tend to over-forecast, especially for longer forecast horizons. A parameter that ‘dampens’ the trend to a flat line in the future was introduced by [8], and is expressed as:

\[
A_t = \alpha X_t + (1-\alpha)(A_{t-1} + \theta T_{t-1})
\]
\[
T_t = \beta (A_t - A_{t-1}) + (1-\beta) \theta T_{t-1}
\]

where:
- \(A_t\) = the attenuated average time series after observing \(X_t\)
- \(T_t\) = the trend estimate
- \(X_t\) = observed values
- \(\alpha, \beta\) = attenuation constants
- \(\theta\) = the damping parameter

RMSE (Eq. 5) and MAPE (Eq. 6) were used for the goodness of fit.

\[
RMSE = \sqrt{\frac{\sum_{t=1}^{n} (A_t - X_t)^2}{n}}
\]

\[
MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{X_t - A_t}{X_t} \right| \times 100
\]

Thus, the Simple model is appropriate for a series in which there are no trends or seasonality because the only smoothing parameter is level. The Holt model is appropriate for a series in which there is a linear trend and no seasonality; its smoothing parameters are level and trend, and are not constrained by each other’s values. The Brown model is appropriate for a series in which there is a linear trend and no seasonality; its smoothing parameters are level and trend, and assumed to be equal. The Damped Trend model is appropriate for a series with a linear trend that is dying out and with no seasonality; its smoothing parameters are level, trend, and damping trend [7].

2.3. Stochastic forecasting

ARIMA (p, d, q) models were used for the stochastic forecast of the seven financial series. These models were tested with multiple combinations for the autoregressive process, moving averages, and transformations to achieve a stationary condition for each one of the series. Equation (7) represents the ARIMA (p, d, q) models.

\[
Y_t = -(\Delta^d Y_t - Y_t) + \phi_0 + \sum_{i=1}^{p} \phi_i \Delta^d Y_{t-i} - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t
\]
\[
\Delta Y = Y_t - Y_{t-1}
\]

where:
- \(Y_t\) = observed values
- \(d\) = differences required for convergence
- \(\Phi\) = autoregressive parameters
- \(\theta\) = moving averages parameters
- \(\phi_0\) = model constant is assumed different from zero
- \(\varepsilon_t\) = error
2.4. Result for deterministic and stochastic forecasts

Table 2 provides results for the various deterministic forecasting methods. The goodness of fit includes RMSE and MAPE as well as the coefficient of determination, R-squared. The column ‘Model Type’ indicates which method proves the best goodness of fit.

<table>
<thead>
<tr>
<th>Table 2. Results for the deterministic forecast.</th>
</tr>
</thead>
<tbody>
<tr>
<td>****</td>
</tr>
<tr>
<td>Driver License Fees</td>
</tr>
<tr>
<td>Federal Aid</td>
</tr>
<tr>
<td>Gas Tax</td>
</tr>
<tr>
<td>Motor Carrier Fees</td>
</tr>
<tr>
<td>Registration Fees</td>
</tr>
<tr>
<td>Special Fuel Tax</td>
</tr>
<tr>
<td>Total State Revenue</td>
</tr>
</tbody>
</table>

Thus, the Simple method and Holt method both provided adequate forecasts for two cases each, and Brown method provided adequate forecasts for three cases.

Table 3 shows the results associated with the application of different stochastic processes used with each of the seven financial series.

<table>
<thead>
<tr>
<th>Table 3. Results of the stochastic forecast.</th>
</tr>
</thead>
<tbody>
<tr>
<td>****</td>
</tr>
<tr>
<td>Driver License Fees</td>
</tr>
<tr>
<td>Federal Aid</td>
</tr>
<tr>
<td>Gas Tax</td>
</tr>
<tr>
<td>Motor Carrier Fees</td>
</tr>
<tr>
<td>Registration Fees</td>
</tr>
<tr>
<td>Special Fuel Tax</td>
</tr>
<tr>
<td>Total State Revenue</td>
</tr>
</tbody>
</table>

Table 2 shows that the best stochastic representation for each of the series was an ARIMA (0, 1, 0). Nevertheless, no ARIMA process (p, d, q) provided the best model to predict each of the series. The goodness of fit from Tables 2 and 3 shows that for each of the seven series, deterministic methods are better than stochastic processes for the study data.

2.5. Forecasts with functional forms

In addition to deterministic and stochastic processes used to forecast financial series, classical functional forms (Linear, Logarithmic, Inverse, Quadratic, Cubic, Growth-Exponential, and Logistic) were used. The coefficient of determination (R-squared) was used as the goodness of fit, as shown in Table 4.
Table 4. Goodness of fit for various functional forms.

<table>
<thead>
<tr>
<th></th>
<th>Linear</th>
<th>Logarithmic</th>
<th>Inverse</th>
<th>Quadratic</th>
<th>Cubic</th>
<th>Grow-Exponential</th>
<th>Logistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Federal Aid</td>
<td>0.752</td>
<td>0.679</td>
<td>0.424</td>
<td>0.767</td>
<td>0.859</td>
<td>0.808</td>
<td>0.808</td>
</tr>
<tr>
<td>Driver License Fees</td>
<td>0.844</td>
<td>0.612</td>
<td>0.326</td>
<td>0.937</td>
<td>0.941</td>
<td>0.887</td>
<td>0.887</td>
</tr>
<tr>
<td>Gas Tax</td>
<td>0.225</td>
<td>0.492</td>
<td>0.636</td>
<td>0.779</td>
<td>0.897</td>
<td>0.241</td>
<td>0.241</td>
</tr>
<tr>
<td>Motor Carrier Fees</td>
<td>0.270</td>
<td>0.410</td>
<td>0.383</td>
<td>0.561</td>
<td>0.572</td>
<td>0.300</td>
<td>0.300</td>
</tr>
<tr>
<td>Registration Fees</td>
<td>0.710</td>
<td>0.889</td>
<td>0.819</td>
<td>0.893</td>
<td>0.938</td>
<td>0.693</td>
<td>0.693</td>
</tr>
<tr>
<td>Special Fuel Tax</td>
<td>0.032</td>
<td>0.161</td>
<td>0.263</td>
<td>0.549</td>
<td>0.656</td>
<td>0.045</td>
<td>0.045</td>
</tr>
<tr>
<td>Total State Revenue</td>
<td>0.394</td>
<td>0.630</td>
<td>0.670</td>
<td>0.759</td>
<td>0.842</td>
<td>0.414</td>
<td>0.414</td>
</tr>
</tbody>
</table>

Table 4 shows that for all financial series, the best fit was obtained using the Cubic form, with an average level adjustment of 82%. Figure 1 and 2 illustrate the various levels of adjustments.

3. Conclusions

This research aimed to forecast various financial data series based on data received from the Nevada Department of Transportation. The forecasts are expected to provide superior data of future revenue compared to the existing methods currently in practice. From the perspective of a state Department of Transportation, of significant importance are the projected state gas tax, special fuel tax, and federal aid revenues.

This research provided an opportunity to choose the best-fit model in conjunction with familiarity with the data and the local and state-wide planned economic conditions. Further studies deemed useful include applying additional variables to the models in order to capture the effect of those variables on forecasted revenue for the next five-year and 10-year horizons. The Logistic Growth function provided a better fit for projecting federal aid revenue over the five-year horizon. The Cubic function led to more reliable revenue projections for the state gas tax, whereas the Brown model provided good forecasts for the total highway-fund revenues.

Historically, straight-line projections used by public sector agencies have not been in line with actual data. In most cases, such straight-line projections have led to grossly overestimating or grossly underestimating the revenues. It is imperative for the decision-makers and system operators to have access to more accurate and robust forecast models that capture the various historical shifts in data to enable them to plan, design, build, and expand the system in line with changing demographics and economic realities. In addition, it will help them be prepared for unusual circumstances and economic downturns.

In this research, deterministic and stochastic methods were tested. The best model specification was obtained using RMSE and MAPE as the goodness of fit. In the context of deterministic methods, the Simple method and Holt method provided adequate forecasts for two cases each, while Brown method provided adequate forecasts for the remaining three cases. The stochastic process, using the ARIMA method, did not find an acceptable goodness of fit for any of the cases. The absence of large datasets is likely to preclude an appropriate estimation using the ARIMA (p, d, q) method [9].

In addition, tests were performed using functional curves, including Linear, Logarithmic, Inverse, Quadratic, Cubic, Growth-Exponential and Logistic. The best fits were obtained using the Cubic functional form, with an average coefficient of determination of 82%.
Fig. 1. Model fit for all the financial series.
Fig. 2. Model fit for all the financial series, continued.

References