FORECASTING TOURIST ARRIVALS TO BALEARIC ISLANDS USING GENETIC PROGRAMMING

Marcos Álvarez-Díaz (malvarezd@cre.sanostra.es)
Josep Mateu-Sbert
Jaume Rosselló-Nadal
Centre de Recerca Econòmica (UIB-Sa Nostra) Carretara Valldemossa km. 7.5
07122-UNIVERSITAT DE LES ILLES BALEARS

ABSTRACT
Traditionally, univariate time-series models have largely dominated forecasting for international tourism demand. In this paper, the ability of a Genetic Program (GP) to predict monthly tourist arrivals from UK and Germany to Balearic Islands (Spain) is explored. GP has already been employed satisfactorily in different scientific areas, including economics. The technique shows different advantages regarding to other forecasting methods. Firstly, it does not assume a priori a rigid functional form of the model. Secondly, it is more robust and easy-to-use than other non-parametric methods. Finally, it provides explicitly a mathematical equation which allows a simple ad hoc interpretation of the results. Comparing the performance of the proposed technique against other method commonly used in tourism forecasting (no-change model, Moving Average and ARIMA), the empirical results reveal that GP can be a valuable tool in this field.

Palabras claves:
Genetic Programming; Tourism Forecasting; Diebold-Mariano Test.

Clasificación JEL (Journal Economic Literature):
C02;C53;L83

Área temática: 2
1. INTRODUCTION

Given the perishable nature of tourism services, there exists an important need to obtain accurate forecasts of future business activity (Archer, 1987; Athiyaman & Robertson, 1992). Certainly, forecasting plays a crucial role in tourism planning both in the short and the long run. However, from a merely practical point of view, tourism industry is much more interested in getting good predictions in the short-term. Needs in the hospitality, transport and accommodation sectors have become more short-term in focus, and they can change rapidly with changing market demand. Therefore, increasing the accuracy of short-term forecasts is an essential requirement to improve the managerial, operational, and tactical decision-making process especially in the private sector.

Identifying an optimal forecasting model for tourism demand has received considerable interest in the specialized literature. Traditionally, the different modeling approaches can be classified into two categories: Multivariate and Univariate Time Series Models. Nevertheless, assuming a parametric point of view might cause problems of over-parameterizing or misspecification, which leads to a lack of generalisation of the model in both cases and, in consequence, a small predictive ability when new observations are considered (Bishop, 1995).

Relatively recently, new advances carried out in Computer Science allowed developing, improving and applying sophisticated non-parametric techniques for the estimation and prediction of different phenomena, including tourism time series data. The goal has been to discover and exploit hidden nonlinear patterns using methods which permit to obtain a model without imposing any a-priori and discrentional assumption on its functional form. In the specific case of tourism forecasting, Artificial Neural Networks (ANN) have been widely applied in the last years to predict tourist arrivals (Law & Au, 1999; Burger et al., 2001; Cho, 2003). Other novel forecasting methods have been used in order to discover a possible predictive improvement. For example, Wang (2003) applied a nonparametric method based on fuzzy premises, and Pai et al. (2006) and Chen & Wang (2007) introduced Support Vector Machine (SVM) models into the tourist predictive problem. In this study a novel predictive tool called Genetic Programming (GP) is applied. The technique, developed in Computer Science and inspired on a natural process, has already been employed satisfactorily in different scientific areas, including Economics (Koza, 1995; Beenstock & Szpiro, 2002) and Finance (Kaboudan, 2000). Chen & Wang (2007) used a
Genetic Algorithm to search optimal parameters of a SVM to predict tourist arrivals to China but, to the authors’ knowledge, the present paper constitutes the first time that a GP is directly applied in tourism forecasting. This method offers multitude of advantages regarding to other forecasting techniques. Firstly, GP does not have any initial restriction on the functional form underlying in the data. Secondly, GP is more robust and easy-to-use than other non-parametric methods like ANN. Finally, it provides explicitly a mathematical equation which allows a simple \textit{ad hoc} interpretation of the results. However, as opposed to these advantages, the technique usually has the difficulty of being computationally intensive and requires a considerable number of observations in order to obtain an accurate result.

The main purpose of this paper is to examine the forecasting accuracy of a genetic programming versus other traditional univariate modeling approaches such as no-change model (NC), Moving Average (MA) and ARIMA in the context of demand for international tourism in the Balearic Islands (Spain). Tourism constitutes the main and fundamental economic activity in this zone where more than 60\% of the Balearic GDP is directly or indirectly related to the tourism. Because of the obvious importance of tourism in this local economy, studies on forecasting the number of arrivals are considered of great relevance. Specifically, the analysis is centered in predicting the tourist arrivals from Germany and United Kingdom (UK). This decision is justified because of the predominance of these two nationalities in the Balearic tourism (almost 80\% of the tourist arrivals) and, secondly, because for both cases they are time series with enough entries to apply the proposed method.

2. GENETIC PROGRAMMING

Genetic Algorithms, originally developed by Holland (1975) and later spread by Goldberg (1989) and Mitchell (2001), enclose a whole series of computing procedures inspired in biological concepts based on the Theory of Evolution of Species: survival of the fittest individuals, reproduction, and birth of offspring with a good genetic heritage. The basic characteristic of these procedures is to use some evolutionary rules observed in Nature as inspiration for solving certain mathematical optimization process. Specifically, from the evolution of a random set of possible solutions and by means of applying operators based on natural selection concepts, these methods allow finding a good approximation to the solution of different optimization problems, including modeling issues.
In the specialized literature there is not a definition of genetic algorithms commonly accepted which allows distinguishing them from other computational evolutionary methods. However, there exist many programs considered as genetic algorithms which present the following common elements: initial population of possible solutions to the problem, selection process using some fit criterion, and use of crossover and random mutation to generate new solutions (Mitchell, 2001). Different variations of genetic algorithms have been applied to a large number of scientific and engineering problems. In this paper, a kind of genetic algorithm called genetic programming (Koza, 1992; Álvarez et al., 2001) is used as a tool to predict the tourist arrivals to Balearic Islands. Therefore, the main goal is to find an optimal predictive model among a collection of multiple candidate equations. This searching process will be done replicating in a computer some basic evolutionary rules. The artificial evolutionary process was programmed in FORTRAN, and it can be explained by means of a series of simple stages. At a first stage, the genetic programming creates a random initial population of N mathematical equations susceptible of representing accurately the time series evolution. These mathematical equations are created by means of a random combination of operators and arguments in the following way:

\[ S_j : ((A \otimes B) \otimes (C \otimes D)) \quad \forall 1 \leq j \leq N \]

where A, B, C, and D are the arguments, the symbol \( \otimes \) represents the mathematical operators and the subscript \( j \) refers to each one of the N equations belonging to the initial population. These arguments can be real numbers included in a certain interval (the equation coefficients) or independent variables (delays of the variable). Besides, the mathematical operators \( (\otimes) \) used will be sum (+), subtraction (-), multiplication (\( \cdot \)) and division (\( / \)), being the latter ‘protected’ to prevent zero divisors. It is also possible to include other mathematical operators (such as logarithm or the trigonometric ones) but at the expense of increasing the complexity in the functional optimisation process. Moreover, previous studies on genetic programming have demonstrated that it is possible to describe complex dynamics with mathematical expressions that are built simply with these four arithmetical operators (Szpiro, 1997; Yadavalli et al., 1999; Álvarez et al., 2001).

At a second stage, after determining the initial population of candidates, the evolution process starts selecting those equations that fit best to the problem. For this purpose, the R-Square has been adopted as fitness criterion. This performance measure is defined as:
\[ R^2_j = 1 - \frac{\sum_{i=1}^{T} (A_i - \hat{A}_i)^2}{\sum_{i=1}^{T} (A_i - \text{mean}(A_i))^2} \quad \forall 1 \leq j \leq N \]

where \( R^2_j \) is the R-Square obtained by equation \( j \), \( A_i \) is the observed value, \( \hat{A}_i \) is the predicted value and \( T \) is the total number of observations in the sub-sample employed to train the genetic program. Later on, all equations of the initial population are classified in decreasing order according to the value of \( R^2_j \). Those equations whose value of \( R^2_j \) is very low are rejected, while those with a high value are more likely to survive, being the base for the next generation of equations.

The equations that survived after the selection process are used to create the equations of a new generation (i.e., reproduction process). In order to do that the so-called genetic operators will be applied: cloning, crossover and mutation. With the cloning operator, the fittest equations are replicated in the next generation. With the crossover operator pairs of equations with high values of \( R^2_j \) are selected and they exchange part of their arguments and of their mathematical operators. Finally, mutation means that any operator or argument is randomly replaced in a small number of equations. The first top ranked individuals are exempted from mutation, so that their information is not lost. Let us consider, for example, that the following equations belong to the initial population:

\[ S_1 : (A + B)/C \]
\[ S_2 : (D \cdot E) - G \]

where A, B, C, D, E and G are the equation arguments (coefficients and time series delays). Suppose that both expressions will survive the selection process and so they become the base equations for the next generation. The crossover operator means the random selection of a block of operators and arguments in each equation and their later exchange. For instance, let us suppose that the block \( (A+B) \) in expression \( S_1 \) and the argument \( G \) in expression \( S_2 \) have been selected. By means of an exchange of blocks two new equations appear as follows:

\[ S_3 : G / C \]
\[ S_4 : (D \cdot E) - (A + B) \]
As one can observe, the new equations inherit certain features from their parents. Now let us suppose that the expression $S_1$ is selected again and the mutation operator is applied. So, the following equation can be obtained from $S_1$:

$$S_5 : (A \cdot B) / C$$

where the mutation was the random alteration of a mathematical operator.

In short, the new population created from the initial population of equations is composed of cloned equations (such as $S_2$), mutated expressions (such as $S_3$), or crossed (such as $S_3$ and $S_4$). From this moment, the process will repeat the selection and reproduction stages in an iterative way. After a given number of generations, determined by the user, the iteration procedure ceases and an optimal mapping $\hat{A}_t = F(A_{t-1}, A_{t-2}, ..., A_{t-m})$ is given by the strongest mathematical equation in the final population.

3-. EMPIRICAL RESULTS

Monthly tourist arrivals to Balearic airports from Germany and UK were obtained from the yearly reports of CITTIB (tourist statistical center belonging to the Balearic government), and available under request at www.inestur.es. The period of collected data goes from January 1980 to December 2006. Therefore, the final sample comprises a total of 324 observations to carry out the predictive exercise.

Following the technical and practical recommendations proposed in the literature on Computer Science (Bishop, 1995) and tourism forecasting (Palmer et al., 2006), the sample was divided into three sub-samples. The first one, called training period, is composed by the first 200 observations and it was reserved exclusively for the evolution of the GP. The next 70 observations conform the selection period, which is employed to look for an optimum number of inputs. The inputs will be delays of the variable under study ($A_{t-1}, A_{t-2}, ... , A_{t-m}$), being $m$ the optimal value. Moreover, the selection period is also used to adjust certain technical aspects of the GP setup. Finally, the last sub-sample, called out-of-sample period, contains the remaining 54 observations. The main goal of these untouched observations is to verify the validity of the definitive model. This methodological procedure is necessary to detect possible spurious results and guaranty that the method is capable of generalizing and performing well with new data.

The fitting criterion used to validate the forecasting accuracy in the different periods is the R-Square, defined by expression (2) and where $A_t$ is the monthly tourist arrival to Balearic
Forecasting tourist arrivals to Balearic Islands using genetic programming

Islands from Germany or from UK. The R-Square must be similar in the three sub-samples, and relatively high in the out-of-sample period. If this condition was verified, it would be proved the ability of the model to generalize new observations and, therefore, it would be confirmed the non-existence of overfitting problems and the consistency of the method.

Analyzing one-period-ahead forecasts in the selection period, Figure 1 shows the sensitivity of GP when different delays are considered (specifically, from 2 to 60 delays). As it can be observed, British and German arrivals show a similar forecasting behavior. From delay 1 to 12 the R-Square goes up; and from delay 12 on the fit criterion is stabilized around a constant value (0.98 and 0.90 for the British and German case, respectively). This predictive behavior is because of the strong seasonal effect existing in the analyzed time series. Once the delay 12 is included, additional delays do not involve a strong predictive gain. Nevertheless and following the recommendation suggested by Casdagli (1989), the number of delays which provides the highest R-Square in the selection period was chosen.

**Figure 1. Choosing an Optimum Delay**

![Figure 1](image)

Once the optimum value of delays has been defined for both time series, the GP starts the evolutionary process searching an optimal model. Table 1 depicts the number of delays, the solution equations and the forecasting results for one period ahead. Some interesting comments can be mentioned analysing this table. First of all, the structure of the optimal equations found by the genetic program is similar for both time series. They show a strong dependence regarding to $A_{t-12}$ plus a non-linear component depending weakly on other delays. Secondly, the Out-of-sample R-Square presents a relatively high value and it is similar to the other sub-samples. Therefore, it seems that the models can generalize quite well new observations and, in consequence, the consistency and the absence of overfitting problems are confirmed. Finally, it is
observed how the GP can get much more accurate predictions in the British case than in the German one.

**Table 1. Model and Forecasting Accuracy of GP**

<table>
<thead>
<tr>
<th>Number of Delays</th>
<th>R-Square</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Selection</td>
</tr>
<tr>
<td>British Arrivals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.9756</td>
<td>0.9874</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**German Arrivals**

<table>
<thead>
<tr>
<th>Number of Delays</th>
<th>R-Square</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Selection</td>
</tr>
<tr>
<td>24</td>
<td>0.9657</td>
<td>0.9168</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 gives information about the forecasting results obtained using different traditional predictive methods such as NC model (one or twelve periods delayed), MA and, finally, ARIMA model. For the last two methods, the selected orders were those that maximized the fit criterion in the selection period. In this way and for both time series, the best MA structure was

$$\hat{A}_t = \frac{A_{t-12} + A_{t-24}}{2}$$

and the optimum ARIMA estimated model for British arrivals was

$$\hat{A}_t = A_{t-12} + e^{0.51 \cdot d_{t-1} + 0.28 \cdot d_{t-2} + 0.25 \cdot d_{t-12} - 0.73 \cdot \hat{d}_{t-12} + 0.52 \cdot \hat{d}_{t-12}}$$

and for the German case

$$\hat{A}_t = A_{t-12} + e^{0.15 \cdot d_{t-1} + 0.14 \cdot d_{t-5} - 0.13 \cdot d_{t-7} + 0.16 \cdot d_{t-8} + 0.01 \cdot d_{t-12} + 0.24 \cdot d_{t-12} - 0.46 \cdot d_{t-12} + 0.14 \cdot d_{t-12} - 0.14 \cdot d_{t-5} + 0.13 \cdot d_{t-7} + 0.56 \cdot d_{t-12} + 0.27 \cdot \hat{d}_{t-12}}$$

where $d_t$ is the difference of the tourism arrival logarithm and $\hat{d}_t$ the estimated value. Results show how all forecasting methods get a high and similar out-of-sample R-Square, except the non-change model one period delayed. This result is due to the strong dependence on $A_{t-12}$. Considering exclusively $A_{t-12}$ as predictor, a high out-of-sample R-square value of 0.9777 and 0.9277 is already achieved for the British and German arrivals, respectively.

Table 2 also shows the Diebold and Mariano (1995) test (D-M). This test checks whether the GP forecasts are statistically different from the traditional methods employed in this study. These authors show that, under the null hypothesis of equal forecasts ability between methods, the
statistic follows asymptotically a standard normal distribution. A positive and statistically significant D-M test would imply to reject the null hypothesis and, in consequence, it could be asserted that the GP provides statistically better predictions that the forecasting model used for comparison purposes.

Table 2. Out-Of-Sample Comparison between GP and different traditional methods.

<table>
<thead>
<tr>
<th>METHOD</th>
<th>(a) British Case</th>
<th>(b) German Case</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R-Square D-M Test (p-value)</td>
<td>R-Square D-M Test (p-value)</td>
</tr>
<tr>
<td>No-Change Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{A}<em>t = A</em>{t-1}$</td>
<td>0.6084 0.5708 (0.00)</td>
<td>3.2865 4.18 (0.00)</td>
</tr>
<tr>
<td>$\hat{A}<em>t = A</em>{t-12}$</td>
<td>0.9777 0.9277 (0.10)</td>
<td>1.6410 0.3828 (0.70)</td>
</tr>
<tr>
<td>Moving Average (MA(2))</td>
<td>0.9796 0.1212 (0.90)</td>
<td>0.9104 2.7768 (0.00)</td>
</tr>
<tr>
<td>ARIMA</td>
<td>0.9868 -1.0159 (0.31)</td>
<td>0.9307 0.7809 (0.4349)</td>
</tr>
</tbody>
</table>

Observing Table 2 (a) for the specific case of British arrivals, the D-M test reveals that the solution equation obtained by the GP gets more accurate predictions than the NC model and MA (2), but this improvement is only statistically significant for the NC predictor $\hat{A}_t = A_{t-1}$. On the other hand, the GP produces worse predictions than the ARIMA model, but the D-M test reveals that there not exist statistically differences between the forecasts of both methods.

Table 2 (b) offers information about the predictive comparison between GP and the rest of the proposed models to predict German arrivals. In this case, the GP outperforms all traditional methods; and this predictive improvement is statistically significant in the comparison with NC model and MA. Two possible reasons could explain the GP predictive success. Firstly, the time series is disturbed by the presence of noise and the GP is less sensitive to these distortions. Secondly, the existence of a weak non-linear structure in the time series could be detected and better exploited by a non-parametric method like the GP.

Up to this point the forecasting results one period ahead were analyzed, but it can be also interesting to analyze what happen when more than one period is considered. Only few studies have tried to forecast tourist time series different periods into the future. Burger et al. (2001) analyzed the ability of a neural network to predict 3, 6 and 12 months ahead. They found that the longer the forecasting period, the poorer will be the prediction. However, if the forecast period was exactly 12 months ahead, the ANN performed fairly well. Palmer et al. (2006) also applied an ANN to predict tourism expenditure in the Balearic Islands. They concluded that expanding the
forecast horizon did not lead to a noticeable predictive decrease. In the present application, Figure 2 describes the ability of the GP to predict from one to twelve periods ahead. As we can see, the out-of-sample R-Square is stabilized for the different forecasting horizons around 0.980 in the British case, and 0.942 for the German arrivals. Therefore, the predictive empirical results corroborate those obtained by Palmer et al. (2006).

**Figure 2. Prediction to Different Periods Ahead**

An additional comment can be mentioned about the stable predictive behavior observed in Figure 2 (a) and (b). Considering the methodology introduced by Sugihara and May (1990), the empirical evidence obtained in this study does not support the existence of a chaotic dynamic in these time series. Chaos is characterized by accurate short-term predictions, but this accuracy is lost when increasing the prediction period. Nevertheless, one must be cautious about this comment because more data are needed and more tools must be applied for searching chaos in tourist time series.

**4- CONCLUSION**

During the last years there has been a growing interest to improve the predictability of tourist arrivals applying new and sophisticated methods based on Computer Science such as neural networks, support vector machines or fuzzy logic. Certainly, the specialized literature encourages the incorporation of additional forecasting techniques in order to achieve a forecasting improvement (Burger et al., 2001). In this study, an emerging technique called Genetic Program (GP) was used to predict monthly tourist arrivals from UK and Germany to Balearic Islands.
(Spain). GP has already confirmed its consistency and forecasting ability in different fields, but it is almost completely unknown in tourism forecasting.

This study compared the performance of the proposed method against different predictive univariate models traditionally used in tourism forecasting (no-change model, Moving Average and ARIMA). The results have verified that GP can be a valuable tool to predict tourist arrivals. The method allows analyzing ex post an equation which represents optimally the dynamic of the time series. In this way, the best mathematical expressions for British and German arrivals show a similar structure: a strong dependence regarding to $A_{t-12}$ plus a nonlinear component depending on other delays. Moreover, GP gets relatively good predictions for the German case in comparison with other more conventional methods. Using the Diebold-Mariano test, the proposed method statistically outperforms the no-change models $\hat{A}_{t} = A_{t-1}$ and $\hat{A}_{t} = A_{t-12}$, an optimal moving average (MA (2)) and it only improves marginally the ARIMA predictions. Worse results were obtained for the British predictive exercise. The Genetic Program only obtains statistically better forecasts in comparison with the predictor $\hat{A}_{t} = A_{t-1}$. The comparison regarding to the other models reveals that the GP method cannot statistically improve the predictive accuracy.

In spite of the promising results obtained in this paper, tourist arrival forecasting and GP functioning are still open research avenues where more efforts must be realized. Other novel methods must be applied, new ones must be developed and the already existing must be improved to reach even more accurate predictions. Furthermore, more research needs to be done to detect chaos in tourist time series. Probably, tourism economists should go one step further on this point in future studies because it is necessary to know and understand more about the nature of tourist time series. Finally, the field of Genetic Algorithms, and of evolutionary computing in general, is relatively recent and new advances and applications are expected in the next years. As Wong and Bodnovich (1998) and Kaboudan (2000) pointed out, these techniques will be intensively used in a near future to solve different problems in economics and finance.
5.- REFERENCES