
Munich Personal RePEc Archive

Modelling Chinese Inbound Tourism Arrivals into Christchurch

Peter Fieger and John Rice

University of New England

1 October 2016

Online at https://mpra.ub.uni-muenchen.de/75468/
MPRA Paper No. 75468, posted 7 December 2016 14:26 UTC

Modelling Chinese Inbound Tourism Arrivals into Christchurch

Abstract

New data and modelling approaches are improving the usefulness of internet search data for forecasting inbound tourist arrivals. This short paper provides evidence of the usefulness of Baidu search data in predicting Chinese inbound tourist arrivals into a specific region in New Zealand. It also compares three modelling approaches, finding a Vector Autoregressive approach the most useful.

Introduction

Forecasting visitor arrivals to a destination using web search data is not new, although modelling advances and improved data availability make an assessment of alternative modelling and recently available data options a useful exercise.

Recent research in the inbound tourism forecasting arena has emphasised the importance of both push and pull factors (Zhou-Grundy and Turner, 2014) – push in relation to source country economic factors and pull factors in relation to recipient country infrastructure. Relatively persistent factor conditions relating to weather and natural attractions also play a fundamental role in the emergence of inbound tourism demand.

Developing better forecasting models is thus an important challenge with significant benefits. Increasingly researchers are identifying the information gap that exists between potential visitors and their intended location as a focal area of research. In this regard, in tourism research, search engine data is proving to be of strong value in improving the predictive ability of models that focus only on pull and push factors (Bangwayo-Skeete and Skeete, 2015).

Google search data has proven useful in enhancing predictive models, but Google is not universally available and, due to a variety of historical reasons, is widely unavailable in China. This limits the use of Google search data in predictive models of outbound China tourism demand.

In this paper, we use search data from Baidu, the major Chinese search engine, to develop an approach that compares more traditional univariate time series analysis methods such as ARIMA and Holt-Winters seasonal forecasting with multivariate Vector Autoregression. Our model is augmented with Baidu search engine data relating to Christchurch, New Zealand's third largest city and a main arrival point for international visitors to the South Island.

Our paper thus has three aims: first to evaluate the predictive capacity of all three approaches, second to assess the value of including search data in the multivariate model and third to assess the value of Baidu search data for predicting Chinese outbound tourism demand.

Empirical Models and Results

This research utilises three different approaches to forecasting visitor arrivals and the aim is to evaluate whether there is some agreement between these forecasting results using presently available information.

The first univariate approach employs the Holt-Winters forecasting method. This method was specifically designed to forecast univariate time series that exhibit a clear seasonal component. The underlying concept of this method is to draw on trend and seasonality

inherent in the time series to produce forecasts. For an in depth description of this approach see Holt (2004) and Winters (1960). As the visitor arrivals time series displays a clearly increasing pattern of seasonality we employed the multiplicative version of the Holt Winters forecasting method.

The second univariate method utilises the autoregressive integrated moving average (ARIMA) approach as devised by Box and Jenkins (1968). This methodology requires the existing times series (e.g. arrival) to be stationary (e.g. the means and standard deviations should be similar at different intervals of the series). After inspecting the visitor arrival time series for existing auto-correlations and partial autocorrelations we decided to build an ARIMA model of the form of (0,1,1)(0,1,1)x12, which mimics the structure of the well-known 'Airline Model'.

Finally, we specified a vector autoregressive model to forecast visitor arrivals. This methodology has emerged relatively recently and was originally pioneered by Watson (1994) and Hamilton (1994). Calibration of a VAR model is significantly more complex than univariate models and requires a number of pre and post model building tests. We utilised the visitor arrival series as well as the Baidu keyword search series for the Chinese terms for 'Christchurch' and 'New Zealand' and 'Queenstown'. Additionally, to account for the seasonality in the data we added the month as an exogenous term. In preparation, all four endogenous time series were transformed using their natural logarithm and were differenced to achieve stationarity. The necessary lag order for this particular model was determined to be 4, using Akaike's Information Criterion. The model showed a high degree of predictive capacity and post-model building analysis revealed that the model exhibited satisfactory statistical properties e.g. there was no remaining autocorrelation at the chosen lag order and that disturbances were largely normally distributed.

The prediction and forecasting results of all three models were superimposed over the original visitor arrival times series and can be seen in Figure 5. On the left of the red line it is obvious that all three models produce predictions that follow the general pattern of the time series very well. We calculated the relative mean square error (RMSE) for each of the three model predictions via

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}}$$

We found that of the three forecasting models used, the two univariate models tended to produce inferior results to the vector autoregressive model model (Holt Winters provided an RMSE of 601.353, ARIMA provided an RMSE of 854.222 and our Vector Autoregressive (VAR) model delivered an RMSE of 443.603).

While prediction of past arrival data is no guarantee for the accurate forecast of future arrivals, it is the VAR model that may be most instructive for forecasting purposes (Figure 5). In the forecasting period, to the right of the red line, we can see a sustained rise in visitor numbers along with a solidified seasonal pattern. The ARIMA and Holt Winters models appear to be slightly more optimistic than the VAR model in their forecasts.

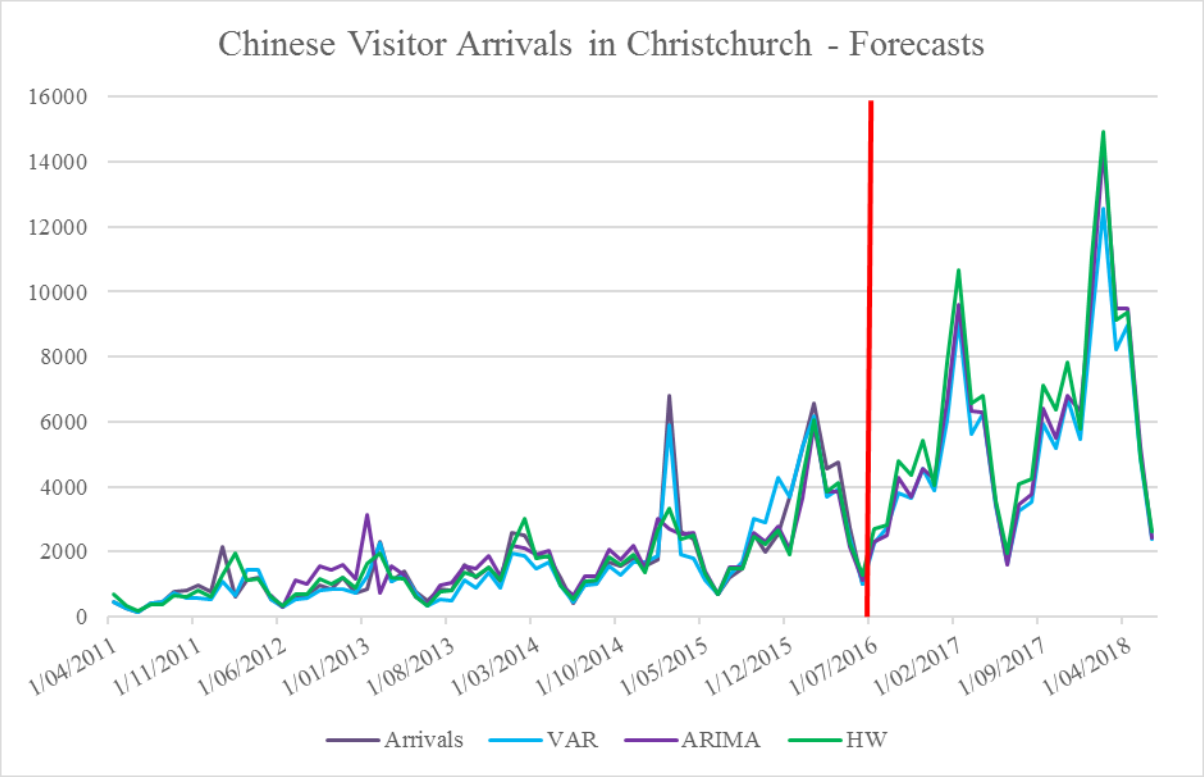

Figure 1: Predictions and forecasts of visitor arrivals

The forecast of the VAR model is of particular interest as it includes the element of internet search data in its prediction. We performed a Granger causality test on our VAR model in an effort to determine whether the endogenous variables in this model are indeed useful for forecasting visitor arrivals. The null hypothesis in this test is that the tested variable does not cause the outcome variable. As can be easily seen, all three hypotheses could be rejected and the predictive value of internet search data for visitor arrival can indeed be confirmed (Table 2).

Granger Causation	Chi2	df	P-value
Baidu 'Christchurch' --> Christchurch arrivals	15.45	4	0.0039
Baidu 'New Zealand' --> Christchurch arrivals	11.55	4	0.0210
Baidu 'Queenstown' --> Christchurch arrivals	19.45	4	0.0005

Table 1: Granger causality test

It is also of interest whether the direct impact of internet search activity can be further defined and quantified. To investigate this question, we estimated the impulse response function that measures the effect that the search for the Chinese term for Christchurch has on Christchurch visitor arrivals of Chinese nationals (Figure 2).

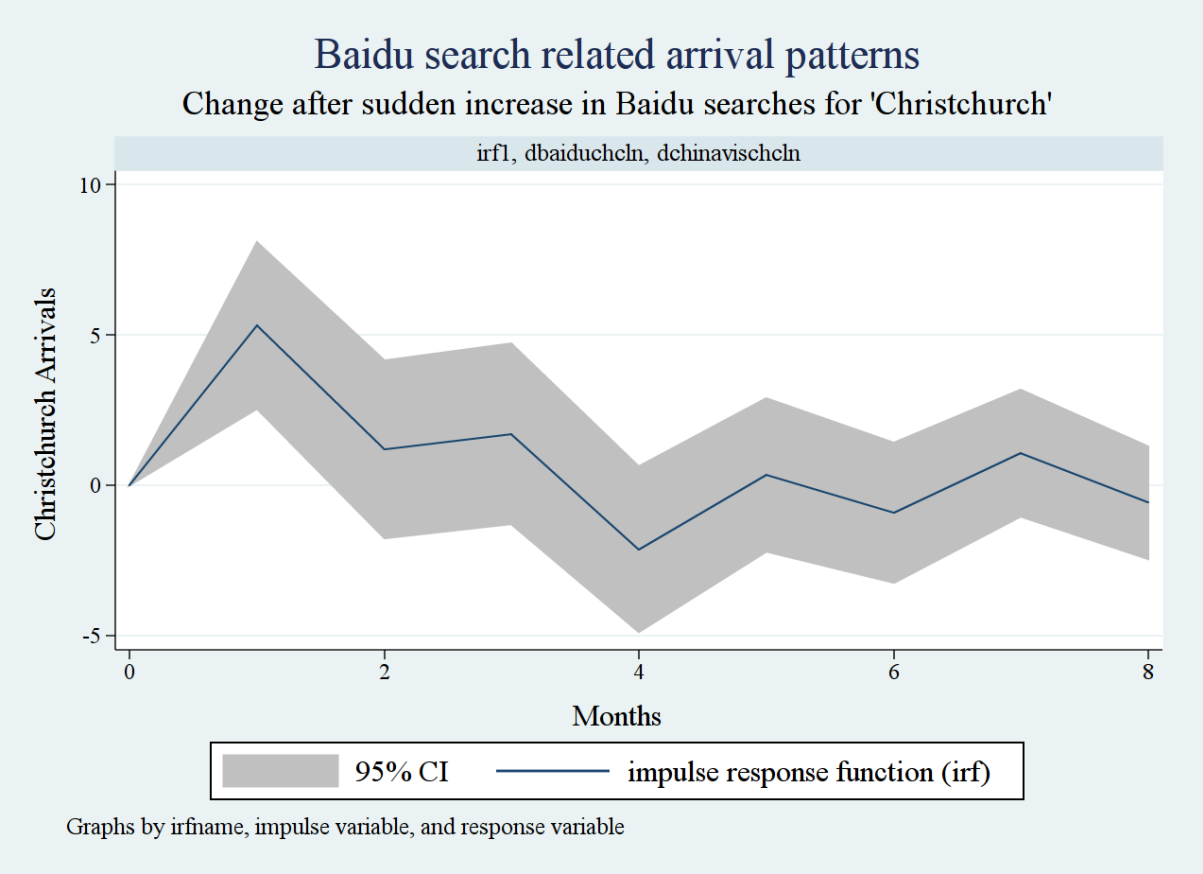

Figure 2: Impulse response function, Christchurch search on visitor arrivals

Figure 2 indicates that surges in search volumes for Christchurch are followed by an immediate significant increase in visitor arrivals in the following month, followed by a continued positive impact (albeit non-significant) for two further months. This finding confirms that internet search data are suitable for supplementing short term visitor forecasting in respect to visitor arrivals originating from China.

Conclusions

Forecasting tourism arrivals from emerging sources is a key challenge for the global tourism industry. In this letter we provide evidence of an important source of new data for forecasting Chinese outbound tourism demand and we provide some guidance on the optimal forecasting model to use in the assessment of this data for the provision of forecasts.

References

Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, *46*, 454-464.

Box, G. & Jenkins, G. (1968). Some recent advances in forecasting and control. *Applied Statistics*, 17(2), 91-109.

Hamilton, J. D., (1994*) Time Series Analysis*, Princeton, Princeton University Press.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1), 5-10.

Watson, M. W., (1994), Vector Autogregression and Co-integration. In Vol. IV of *Handbook of Econometrics*, ed. R.F. Engle and D.L. McFadden. Amsterdam: Elsevier.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, *6*(3), 324-342.

Zhou-Grundy, Y., & Turner, L. W. (2014). The Challenge of Regional Tourism Demand Forecasting: The Case of China. *Journal of Travel Research*, DOI: 0047287513516197.