Annals of Tourism Research 61 (2016) 199–212

Contents lists available at ScienceDirect

Annals of Tourism Research

journal homepage: www.elsevier.com/locate/atoures

Forecasting city arrivals with Google Analytics

Ulrich Gunter *, Irem Önder

MODUL University Vienna, Department of Tourism and Service Management, Am Kahlenberg 1, A-1190 Vienna, Austria

ARTICLE INFO

Article history:
Received 29 December 2015
Revised 5 October 2016
Accepted 10 October 2016

Coordinating editor: Haiyan Song

Keywords:
Bayesian analysis
Big data
City tourism
Factor analysis
Forecast combination
Vector autoregression

ABSTRACT

The ability of 10 Google Analytics website traffic indicators from the Viennese DMO website to predict actual tourist arrivals to Vienna is investigated within the VAR model class. To prevent overparameterization, big data shrinkage methods are applied: Bayesian estimation of the VAR, reduction to a factor-augmented VAR, and application of Bayesian estimation to the FAVAR, the novel Bayesian FAVAR. Forecast accuracy results show that for shorter horizons ($h = 1, 2$ months ahead) a univariate benchmark performs best, while for longer horizons ($h = 3, 6, 12$) forecast combination methods that include the predictive information of Google Analytics perform best, notably combined forecasts based on Bates–Granger weights, on forecast encompassing tests, and on a novel fusion of these two.

Introduction

There has been dramatic growth in Internet usage, especially for travel information search purposes. Each time an individual uses a website he or she leaves traces on that site. These traces can be collected and used for different purposes such as tracking user behavior, recommending products to the customer on their next visit to the website and optimizing website usability. Google Analytics accounts make it possible for businesses to collect these traces from their websites. Although many destination management organizations (DMOs) are collecting these types of information from their websites, this information is usually not used for making managerial decisions, but merely by IT departments to enhance website usability. Often, the interpretation of website traffic indicators such as Google Analytics is not clear to DMO managers: such as what it means to the DMO to have one million website visitors. However, website traffic data can be very informative: showing, for instance, from which countries users originate. This information can then be combined to see if there is a correlation between the country of origin of website visitors and the country of origin of the actual arrivals to the destination.

A primary objective of this article is therefore to show how website traffic data can be used by DMO managers to forecast tourism demand. Tourism demand analysis and forecasting is one of the core areas of tourism economic research since tourism demand ultimately is the basis of all business decisions in tourism (Song, Witt, & Li, 2009), which includes the business decisions of DMOs. These business decisions require accurate tourism demand forecasts in order to reduce the many risks that may occur during the decision making process (Frechtling, 2001).

The theory explaining the search behavior of the visitors to the DMOs' websites is called *information foraging theory*, which is derived from behavioral ecology and which is similar to food foraging theories in anthropology. It was developed

* Corresponding author.
E-mail addresses: ulrich.gunter@modul.ac.at (U. Gunter), irem.onder@modul.ac.at (I. Önder).

http://dx.doi.org/10.1016/j.annals.2016.10.007
0160-7383/

by Pirolli and Card (1999) and states that, when possible, the information search process evolves toward maximization of relevant information per unit cost. In the same sense, the optimal information forager would be the one who maximizes the rate of relevant information gained according to the task environment, the profitabilities of different sources, and the cost of finding and accessing them (Pirolli & Card, 1999). Thus, foraging information from a DMO's website comes at a comparably low unit cost. In a tourism demand forecasting context, the information foraged from the DMOs' websites and expressed by various Google Analytics website traffic indicators can therefore constitute an important predictor for actual arrivals to a destination if the foraged information proves relevant. Relevance of this foraged information, in turn, is given if considering it in appropriate forecast models results in comparatively more accurate forecasts.

Information search on search engines such as Google, Yahoo, or Baidu is one of the first steps in planning a vacation, with travel-related searches comprised of city, country, and region names representing approximately 60% of all travel related searches (Jansen, Ciamacca, & Spink, 2008). From the search engine results, the DMOs' websites can typically be found on the first page of search results and thus potential travelers are immediately directed to the DMOs' websites. Google Analytics in this case would provide researchers with quantitative measures for potential travelers to the destination that can be used for in-depth analysis.

Google Analytics is a free service by Google to the owner of a website that provides website traffic data including all user behavior on the website, such as the number of visitors, the time spent on the website, the number of actions taken on the site, where the users come from, etc. Google Analytics sends the website traffic data to the analytics server by means of a snippet (tracking code) that is included on the website and activated when a visitor views a page on somebody's website (Boswell, 2011). Overall, Google Analytics has a nearly 83% share of the website tracking tools market (W3Techs, 2015), yet this service is generally used only for website quality control and to enhance website user experience.

Some examples of previous studies that have used Google Analytics include measuring the website performance of a cultural tourism website (Plaza, 2011), measuring food composition website visitor statistics (Pakkala, Presser, & Christensen, 2012), as well as understanding Google Analytics data and using it as a communications tool (Kent, Carr, Husted, & Pop, 2011). The contribution of the present study is to investigate whether Google Analytics data for a DMO's website possess predictive information that helps to improve the accuracy of tourism demand forecasts in terms of actual tourist arrivals to a destination.

There is one further study employing Google Analytics for forecasting tourism demand (Yang, Pan, & Song, 2014). However, that study only uses numbers of website visitors and website visits as website traffic indicators, whereas the present study employs 10 Google Analytics indicators in a more comprehensive forecast modeling exercise. Other than that, there are some instructions and codes on the Internet outside the tourism discipline on how to retrieve Google Analytics data through a website's application programming interface (API) and how to forecast the retrieved Google Analytics indicators as measures of website traffic. Typically, univariate forecast models are employed using statistical software such as R (see e.g. Breña Moral, 2007; Granowitz, 2014; Vadera, 2013; Zwitch, 2013).

To this end, the website of the DMO of Vienna, the Vienna Tourist Board, has been selected for scrutiny: www.wien.info. Since Google Analytics data are not open to the public, the authors received permission from the Vienna Tourist Board to use their Google Analytics data for the time period from 2008M08 until 2014M10 for this research. Apart from issues of data availability, Vienna is one of the top-10 city destinations in Europe, with more than 13 million bednights in 2013, followed by Munich, Hamburg and Amsterdam (European Cities Marketing (ECM) & MODUL University Vienna (MU), 2014), and therefore of particular relevance for such a study in the thriving strand of literature on city tourism demand forecasting.

There exist more than 20 Google Analytics indicators, 10 of which were used in the present study based on *data availability* and *potential predictive power*. Given the high dimensionality of the sample (relatively few observations in combination with relatively many variables: total tourist arrivals to Vienna plus the 10 Google Analytics website traffic indicators), methods of big data shrinking become necessary to preclude inaccurate forecasts. Big data in this sense are defined as "*data sets and analytical techniques in applications that are so large (from terabytes to exabytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis, and visualization technologies*" (Chen, Chiang, & Storey, 2012, p. 1166); Google Analytics data clearly fulfill this definition.

A number of articles addressing the predictive performance of different types of big data for tourism demand forecasting have been published, typically underlining the usefulness of the employed types of big data for improvements in forecast accuracy. Gawlik, Kabaria, and Kaur (2011), for instance, use web search volume histories to predict visitor numbers. Jackman and Natiram (2015) examine the usefulness of Google Trends for predicting tourist flows to Barbados using support vector regressions (SVRs). Bangwayo-Skeete and Skeete (2015) apply a mixed-frequency approach using Google Trends to predict tourist arrivals to Caribbean islands. In a study by Önder and Gunter (2016), in which Google Trends web and image search indices are used for English and native language searches, the usefulness of Google Trends for forecasting is confirmed, particularly with respect to native language searches.

Yang, Pan, Evans, and Lv (2015) compare search engine queries from Google and Baidu to predict visitor numbers to Hainan, China, with the result that search queries from Baidu outperform those from Google. Moreover, the Baidu index has been used in a recent article that investigates the tourist flows to the Forbidden City in Beijing, China, concluding that there is a positive impact of employing the Baidu index as an explanatory variable on the accuracy of visitor numbers forecasts (Huang, Zhang, & Ding, 2016).

Since Block Granger causality tests suggest a mutual causality structure between the 11 variables, vector autoregression (VAR, Lim & McAleer, 2001; Oh, 2005; Sims, 1980; Shan & Wilson, 2001; Song & Witt, 2006) and big data shrinking techniques geared towards this model class as suggested in the literature are of interest (Bańbura, Giannone, & Reichlin, 2010; Giannone, Lenza, & Primiceri, 2015): Bayesian estimation of the VAR, resulting in the so-called Bayesian vector autoregression (BVAR, Doan, Litterman, & Sims, 1984; Gunter & Önder, 2015; Song, Smeral, Li, & Chen, 2013; Wong, Song, & Chon, 2006), extraction of common factors from the 10 Google Analytics website traffic indicators using principal component analysis, resulting in the so-called factor-augmented vector autoregression (FAVAR, Bernanke, Boivin, & Eliasz, 2005; New Zealand Institute of Economic Research (NZIER), 2012; Stock & Watson, 1998; Stock & Watson, 2002), as well as a novel combination of these two big data shrinkage methods, resulting in the so-called Bayesian factor-augmented vector autoregression (BFAVAR, Amir-Ahmadi & Uhlig, 2013; Bekiros & Paccagnini, 2014), which has not been employed in tourism demand forecasting so far.

As univariate benchmarks, members of the autoregressive integrated moving average (ARIMA, Box & Jenkins, 1970; Kulendran & Witt, 2001; Li, Song, & Witt, 2006; Song, Romilly, & Liu, 2000; Witt, Song, & Louvieris, 2003), the Error-Trend-Seasonal or ExponenTial Smoothing (ETS, Athanasopoulos, Hyndman, Song, & Wu, 2011; Hyndman, Koehler, Snyder, & Grose, 2002; Hyndman, Koehler, Ord, & Snyder, 2008), and the naïve model classes are employed.

Apart from the single forecast models, also the benefits of multiple forecast combination methods are evaluated. Bates and Granger (1969) indicate that combination forecasts can yield lower forecasting errors, a finding which was later confirmed by Clemen (1989). In addition to two forecast combination methods that have become somewhat standard in the tourism demand forecasting literature, simple uniform and Bates–Granger weights (e.g. Andrawis, Atiya, & El-Shishiny, 2011; Cang, 2011; Chan, Witt, Lee, & Song, 2010; Shen, Li, & Song, 2011; Song, Witt, Wong, & Wu, 2009; Wong, Song, Witt, & Wu, 2007; Wu & Zhang, 2014), two less common but potentially quite powerful forecast combination methods are suggested: uniformly weighted forecast combination based on forecast encompassing tests (Harvey & Newbold, 2000; Shen et al., 2011) and forecast combination based on forecast encompassing tests, where Bates–Granger weights are employed instead of uniform weights. The latter fusion of Bates–Granger weighting and forecast encompassing tests is a novel phenomenon in general (Costantini, Gunter, & Kunst, 2016) and in the tourism demand forecasting literature in particular.

In this study, the ex-ante out-of-sample forecast performance (all variables forecast jointly) of seven single forecast models (VAR(2), BVAR(2), FAVAR(2), BFAVAR(2), MA(2), ETS(*A, N, N*), and Naïve-1) as well as of four forecast combination methods (uniform weights, Bates–Granger weights, forecast encompassing tests with uniform weights, and forecast encompassing tests with Bates–Granger weights) is evaluated in terms of the root mean square error (RMSE) and the mean absolute error (MAE) for forecast horizons $h = 1, 2, 3, 6, 12$ months ahead while using expanding estimation windows (so-called recursive forecasting).

The forecast evaluation results show that for shorter forecast horizons ($h = 1, 2$) the univariate MA(2) model outperforms its competitors in terms of accurately forecasting total tourist arrivals to Vienna. For longer horizons ($h = 3, 6, 12$), however, combined forecasts based on (a) Bates–Granger weights, (b) forecast encompassing tests with uniform weights, and (c) forecast encompassing tests with Bates–Granger weights deliver the lowest RMSE and MAE values, while for $h = 12$ and MAE the single BVAR(2) model outperforms everything else. This underlines the predictive ability of the 10 Google Analytics website traffic indicators representing the information foraged from the website of the DMO of Vienna and the viability of the employed big data shrinking techniques.

The over-parameterized VAR(2) and the Naïve-1 models are often among the worst-performing models and are also significantly outperformed on a regular basis by their competitors in terms of Hansen tests (Hansen, 2005). Therefore, these two forecast models were also often excluded from the forecast combination methods that were subsequently shown to perform best. The good performance of the more sophisticated multivariate forecast models and, in particular, of the more sophisticated forecast combination methods containing this additional information is noteworthy since it is at odds with the traded wisdom that for forecast horizons up to two years univariate forecast models should be expected to produce the most accurate forecasts (Frechtling, 2001).

Data

Tourism demand to Vienna is measured in terms of *Total Arrivals* (domestic and foreign source markets) in all paid forms of accommodation establishment in the greater city area. The monthly tourist arrivals from 2008M08 until 2014M10 are collected from TourMIS (www.tourmis.info), which is a publicly accessible online marketing information system designed for use by tourism managers and researchers.

Website traffic data were retrieved from the Vienna Tourist Board's Google Analytics account. This information is typically used by the Vienna Tourist Board employees to improve their website effectiveness and track users on the site. Since the website traffic data are not publicly available, the authors received permission from the Vienna Tourist Board to access them for the purpose of the current research. The website traffic data retrieved are also restricted to 2008M08 until 2014M10, as the researchers lacked rights to access data beyond 2014M10.

Google Analytics has more than 20 indicators; 10 of which that were selected for this study based on their potential predictive power (see Fig. 1 for a screenshot of a typical Google Analytics session for www.wien.info). The indicators that were excluded were either only sporadically available and/or not directly linked to forecasting such as the type of Internet browser

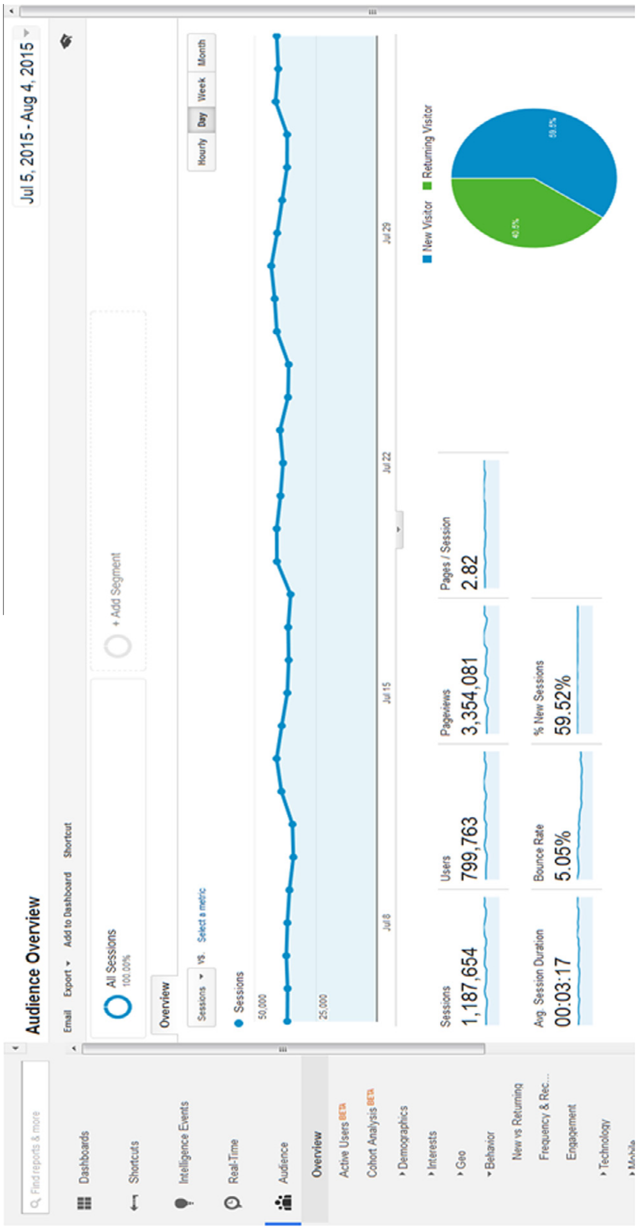

Fig. 1. Screenshot of a Google Analytics session for www.wien.info. Source: Google Analytics data for www.wien.info.

(i.e. Chrome, Firefox, Safari, etc.) used to enter the website. The employed indicators were downloaded as monthly aggregates and included the following:

- *Average Session Duration (in seconds)* shows the engagement of the users on the website based on the time they spend in an average session.
- *Average Time on Page (in seconds)* indicates the engagement of the users based on the time they spend on average on each page.
- *Bounce Rate (in %)* is another online engagement indicator that shows the percentage of single-page sessions such as the sessions, where users left the website without clicking on the links on the page (Google, 2015). This may be a result of bad website design or usability issues, thus this is a negative indicator.
- *New Sessions (in %)* is an estimate of the percentage of first time visits out of all visits to the website.
- *Page Views* is the total number of views of a page that is being tracked by the website analytics tracking code.
- *Returning Visitors (in %)* is an estimate of the percentage of returning visitors out of all visitors to the website.

- *Social Network Referrals* include the total number of referrals that come to the website through social networks such as Facebook or Twitter.
- *Total Sessions* are the total number of sessions, whereby session is defined as the period of time a user is actively engaged with the website.
- *Unique Page Views* are the total number of unique pages viewed.
- *Users* include the total number of website users.

Online engagement is a measure of success for websites as it shows the ability of a website to hold a visitor's attention or encourage them to interact with the website (Meares, 2015). Time spent on a website can be interpreted in two ways: a long session may either indicate that the user cannot find the information he/she is looking for, or it could mean that the user finds the website interesting and engaging, causing them to read the information in detail. In this study the second interpretation is assumed, such that longer sessions are assumed to indicate 'better' websites (read: engaging, fun, or providing useful information). Following this interpretation, Bounce Rate is interpreted negatively since users do not interact with the page.

Except for Average Session Duration and Average Time on Page, which are both given in seconds, the remaining eight Google Analytics website traffic indicators as well as tourist arrivals to Vienna are given in total numbers. This means that the variables Bounce Rate, New Sessions, and Returning Visitors had to be transformed from percentages as retrieved from Google Analytics to total numbers first before they could be used for estimation and forecasting. As can be seen from Fig. 2, all variables are subject to more or less pronounced upward trends (except the Bounce Rate, which is a negative indicator).

Over the period of analysis, the website of the Vienna Tourist Board clearly improved in terms of attracting more visitors, better engaging visitors, and generating more returning visitors. These are all signs that the website provides valuable and relevant information, which induces users to return. Since the usage and the usefulness of the DMO's website have increased over some considerable amount of time, the improvements in the Google Analytics indicators may also be attributed to appropriate updates and redesigns of the website. In addition to the upward trend, tourist arrivals to Vienna are characterized by a distinct seasonal pattern. Therefore, moving average filters are applied to all variables to seasonally adjust the data. Furthermore, to ensure a linear functional relationship, natural logarithms are taken of all variables.

Concerning the trending behavior of the data, which is still present after seasonal adjustment and taking natural logarithms, Augmented Dickey Fuller (ADF) tests for non-seasonal unit roots allowing for linear time trends and intercepts are employed to investigate whether these trends are stochastic or deterministic. Since the null hypothesis of the presence of a non-seasonal unit root cannot be rejected for six of the 11 variables (see Table 1), first differences are taken of all variables to ensure stationarity and to prevent the use of variables with different degrees of integration for estimation and forecasting.[1]

Finally, Block Granger causality tests frequently reject their null hypothesis of excluded variables jointly not Granger causing the respective dependent variable, except for Average Time on Page and Bounce Rate (see Table 2). This implies that employing the 10 Google Analytics website traffic indicators as explanatory variables (or predictors) may be worthwhile in terms of forecasting tourist arrivals to Vienna, thus rendering the use of multivariate or econometric forecast models a viable option in general. However, these test results also imply that the explanatory variables should not be interpreted and treated as exogenous since Granger causality is not monodirectional in the majority of cases. This bidirectional causality is not surprising since also actual visitors in Vienna are likely to use or come back to its DMO's website for (potentially more specific) searches about their destination of choice.

Methodology

Rival forecast models

As a consequence of the block Granger causality test results, multivariate models that relax the assumption of exogeneity of the explanatory variables are employed.[2] A multivariate model which relaxes the assumption of exogeneity of the explanatory variables is the VAR(p) approach favored by Sims (1980). In its general form such a classical VAR(p) reads:

$$X_t = A_t + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \Upsilon_t, \tag{1}$$

where A_t is a $k \times 1$ intercept vector, X_t a $k \times 1$ vector of the observed variables (tourist arrivals to Vienna $\Delta q_t = \ln(Q_t) - \ln(Q_{t-1})$, with Δ denoting the first difference operator, plus the 10 Google Analytics website traffic indicators; all seasonally adjusted, in natural logarithms, and first differenced), Φ a $k \times k$ coefficient matrix, and Υ_t a $k \times 1$ vector of error terms to be i.i.d $\sim N(0, \Sigma_\Upsilon)$ for time points $t = 2009M8, \ldots, 2014M10$. Since the error terms are assumed to be contempora-

[1] Testing for seasonal unit roots in addition to the non-seasonal ones, for instance by applying the monthly Hylleberg-Engle-Granger-Yoo test (HEGY, Beaulieu & Miron, 1993; Hylleberg, Engle, Granger, & Yoo, 1990), was not deemed viable for the present sample because of two reasons. Firstly, the data had already been seasonally adjusted before ADF tests were conducted. Secondly, finding seasonal unit roots in monthly data would have induced seasonal differencing for the transformed variables to remain interpretable. This procedure would have reduced the already quite short sample considerably.

[2] While other multivariate models typically used in tourism demand forecasting as laid out, for instance, in Song et al. (2009) may theoretically be feasible, they are ruled out by the Block Granger causality test results since those rest on the exogeneity assumption of the explanatory variables.

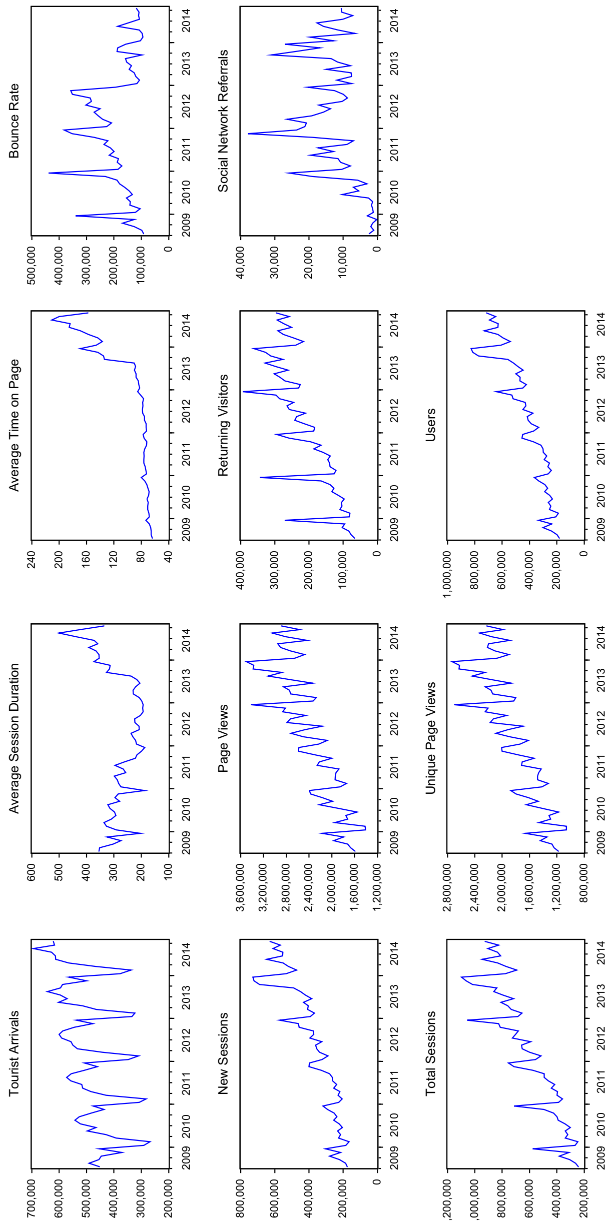

Fig. 2. Evolution of tourist arrivals and 10 Google Analytics website traffic indicators from 2009M7 to 2014M10. Source: www.tourmis.info and Google Analytics data for www.wien.info.

neously but not serially correlated, Eq. (1) can be estimated using ordinary least squares (OLS, e.g. Song, Witt, & Jensen, 2003). The optimal lag order $\hat{p}$ is selected by the Akaike information criterion (AIC) ($\hat{p} = 2$), which can be shown to select $\hat{p}$ of a VAR(p) such that its forecast mean square error is minimized (Lütkepohl, 2005). Since its first introduction (Kulendran & King, 1997; Kulendran & Witt, 1997), the VAR(p) has been frequently employed in tourism demand forecasting (see Lim & McAleer, 2001; Oh, 2005; Shan & Wilson, 2001; Song & Witt, 2006, for some recent examples).

Given the short sample length of only 61 observations for estimation after adjustments for the whole sample ($2009M10 - 2014M10$) and even fewer for forecasting later on, combined with the high dimensionality of the VAR with 12 variables (including the intercept), estimating an unconstrained VAR(2) by OLS or pure maximum likelihood results in estimating an over-parameterized model, which is likely to be detrimental to its forecast performance. Therefore, methods for big data shrinkage are applied to create three rival forecast models to the classical VAR(2) of dimension $k = 12$: estimation of the VAR(2) with Bayesian methods (BVAR), reduction of the dimensionality of the VAR(p) by extracting common factors from the 10 Google Analytics website traffic indicators through principal component analysis (FAVAR), and application of Bayesian methods to the aforementioned FAVAR (BFAVAR).

Table 1
ADF test results.

Variable	Lag order $\hat{p}$	ADF statistic	p-value
Average Session Duration	0	-2.17837	0.4929
Average Time on Page	0	-1.58203	0.7891
Bounce Rate	0	-2.93565	0.1587
New Sessions	0	-4.87526^{***}	0.0010
Page Views	0	-3.86655^{**}	0.0193
Returning Visitors	2	-1.60358	0.7802
Social Network Referrals	0	-3.44035^{*}	0.0552
Total Sessions	2	-0.50914	0.9805
Tourist Arrivals	0	-7.90274^{***}	0.0000
Unique Page Views	2	-0.08141	0.9941
Users	0	-5.22147^{***}	0.0003

Source: TourMIS, Google Analytics data for www.wien.info, and own calculations. All individual ADF tests include an intercept and a linear time trend. (***) denotes statistical significance of the ADF statistics at the 1%, (**) at the 5% and (*) at the 10% level. The optimal lag orders $\hat{p}$ for the ADF tests are determined by the Bayesian information criterion (BIC).

Table 2
Block Granger causality test results.

Dependent variable	Excluded variables	χ^2-statistic	p-value
Average Session Duration	All other	48.18439^{***}	0.0004
Average Time on Page	All other	9.56200	0.9754
Bounce Rate	All other	24.85139	0.2072
New Sessions	All other	43.62744^{***}	0.0017
Page Views	All other	29.18036^{*}	0.0843
Returning Visitors	All other	44.66499^{***}	0.0012
Social Network Referrals	All other	30.91207^{*}	0.0564
Total Sessions	All other	38.49526^{***}	0.0077
Tourist Arrivals	All other	32.27653^{**}	0.0404
Unique Page Views	All other	30.27127^{*}	0.0656
Users	All other	40.91088^{***}	0.0038

Source: TourMIS, Google Analytics data for www.wien.info, and own calculations. (***) denotes statistical significance of the χ^2 statistics at the 1%, (**) at the 5% and (*) at the 10% level. The optimal lag order $\hat{p}$ of the VAR(p) is determined by the Akaike information criterion (AIC) ($\hat{p} = 2$, degrees of freedom: 2).

The first big data shrinkage method applied to an over-parameterized VAR(p) is the use of Bayesian methods rather than OLS. In doing so, the so-called Minnesota or Litterman prior for BVAR(p) estimation and forecasting is employed. The Minnesota prior is a so-called informative prior developed by Doan et al. (1984) and Litterman (1986) on an otherwise unconstrained VAR(p) with intercept, which has remained one of the post popular informative priors in the BVAR literature including some later variations (Giannone et al., 2015).

The main differences to the classical VAR(p) estimated by OLS are that the Minnesota prior assumes that the model parameters (i.e. the intercept vector and the coefficient matrices) are random variables, that all VAR equations are centered around a random walk with drift process, that it imposes restrictions on the more distant variable lags rather than eliminating them (lag decay), and that it includes the possibility of imposing less influence of lags of other variables relative to own lags (relative cross variable weight) (Bańbura et al., 2010). Shrinking the unconstrained VAR(p) to a more parsimonious representation by applying informative priors therefore reduces parameter uncertainty, which, in turn, is generally beneficial to forecast accuracy (Karlsson, 2013). However, using uninformative or diffuse priors such as the flat prior, which is equivalent to the classical VAR(p) estimated by OLS, has been shown not to exhibit the benefits of Bayesian shrinkage (Giannone et al., 2015).

As described in Lütkepohl (2005), it is assumed in Bayesian estimation that non-sample information on a generic parameter vector ψ available prior to estimation is summarized in its prior probability density function (PDF) $g(\psi)$. The sample information on ψ, in turn, is summarized in its sample PDF given by $f(\mathbf{y}|\psi)$, which is algebraically identical to the likelihood function $l(\psi|\mathbf{y})$. Applying Bayes' theorem, the following relation between the prior PDF and the sample PDF can be established, where $f(\mathbf{y})$ denotes the unconditional sample density:

$$g(\psi|\mathbf{y}) = \frac{f(\mathbf{y}|\psi)g(\psi)}{f(\mathbf{y})}. \tag{2}$$

Eq. (2) states that the distribution of ψ conditional on the sample information contained in $\mathbf{y}$ can be summarized by $g(\psi|\mathbf{y})$, which is known as posterior PDF. In other words, the posterior distribution, which contains all information available for the parameter vector ψ, is proportional to the likelihood function times the prior PDF:

$$g(\psi|\mathbf{y}) \propto f(\mathbf{y}|\psi)g(\psi) = l(\psi|\mathbf{y})g(\psi). \tag{3}$$

Since the posterior PDF as stated in Eq. (3) cannot be obtained analytically, assumptions reflecting prior beliefs have to be made about the parameter vector ψ in order to obtain it numerically. For the present case, the parameter vector ψ consists of the model parameters and the three BVAR(p) (here: BVAR(2)) hyperparameters that need to be calibrated: overall tightness (set to 0.1, denoting a relatively tight value as recommended for small BVAR(p) systems), relative cross-variable weight (set to 0.5, reflecting symmetric characteristics of the BVAR(p) model), and lag decay (set to 1, representing linear decay), which are meant to obtain a model that is referred to as a "standard BVAR(p)" in Wong et al. (2006). Since its first usage in Wong et al. (2006), the BVAR(p) has not received as much attention as the classical VAR(p) in the tourism forecasting literature (see Gunter & Önder, 2015; Song et al., 2013, for the few recent examples).

The second big data shrinkage method applied to an over-parameterized VAR(p) is the reduction of dimensionality by extracting common factors from the 10 Google Analytics website traffic indicators. The FAVAR(p) approach was pioneered by Bernanke et al. (2005) and apart from one study (New Zealand Institute of Economic Research (NZIER), 2012), the FAVAR(p) approach has never been applied to tourism demand forecasting. In fact, in a FAVAR(p) context, tourism variables have almost exclusively been used as examples of sectoral economic variables from which factors have been extracted for macroeconomic analysis and forecasting (e.g. Chow & Choy, 2009; Ljubaj, 2012; Pang, 2010; Ribon, 2011). In its general form a FAVAR(p) reads:

$$\Phi(L)\begin{bmatrix} \Delta q_t \\ F_t \end{bmatrix} = \Upsilon_t, \tag{4}$$

where $\Phi(L)$ denotes a lag polynomial of finite order p (with L denoting the lag operator). Δq_t denotes again tourist arrivals to Vienna and F_t the unobserved factors together forming a $k \times 1$ vector of variables (including the intercept). Υ_t again is a $k \times 1$ vector of error terms to be i.i.d $\sim N(0, \Sigma_\Upsilon)$ for time points $t = 2009M8, \ldots, 2014M10$. Since the factors F_t in Eq. (4) are unobserved, principal component analysis with factor rotation (to obtain orthogonal factors) is applied to obtain estimates $\widehat{F}_t$ in a first step (Stock & Watson, 1998, 2002). The Kaiser criterion suggests to extract 2 factor estimates $\widehat{F}_t$ from the 10 Google Analytics website traffic indicators resulting in a FAVAR(p) of dimension $k = 4$ since only two principal component factor estimates possess eigenvalues greater than 1. Finally, the optimal lag order $\hat{p}$ is again selected by AIC ($\hat{p} = 2$) and the resulting FAVAR(2) is estimated using OLS.

Finally, a combination of the two aforementioned big data shrinkage methods is proposed: an application of Bayesian estimation as summarized by Eqs. (2) and (3) to the FAVAR(p) model as described by Eq. (4): the so-called BFAVAR(p) (here: BFAVAR(2) of dimension $k = 4$ with the same calibration for the BVAR(p) hyperparameters as above). This model class has only recently been introduced into the macroeconomic analysis and forecasting literature (e.g. Amir-Ahmadi & Uhlig, 2013; Bekiros & Paccagnini, 2014) and – to the best of the authors' knowledge – has never been applied to tourism demand forecasting. Consequently, VAR(2), BVAR(2), FAVAR(2), and BFAVAR(2) models are employed as rival forecast models.

Univariate benchmarks

Concerning univariate benchmarks, members of the ARIMA, the ETS (Hyndman et al., 2002, 2008), and the naïve model classes are employed. ARIMA and ETS have been chosen since Athanasopoulos et al. (2011) find, based on 366 monthly tourism time series, that SARIMA and ETS models feature a higher predictive accuracy than the seasonal naïve benchmark. As the data employed in this study are seasonally adjusted and first differenced, only ARMA, ETS, and naïve models suited for non-seasonal and stationary data are relevant in this study.

The (S) AR(I) MA model class as proposed by Box and Jenkins (1970) has been employed frequently for tourism demand forecasting. Of the numerous studies published in the past two decades that employ these models (Song et al., 2009), the majority has used them as benchmarks when assessing the forecast accuracy of more complex multivariate models (e.g. Kulendran & Witt, 2001; Li et al., 2006; Song et al., 2000; Witt et al., 2003). A standard ARMA (p, q) reads as follows:

$$\varphi(L)\Delta q_t = \alpha + \vartheta(L)\upsilon_t, \tag{5}$$

where $\varphi(L), \vartheta(L)$ denote lag polynomials of finite orders p and q respectively, α an intercept term, and υ_t denotes an error term to be i.i.d. $\sim N(0, \sigma_\upsilon^2)$. Eq. (5) can be estimated by using OLS. Visual inspection of the autocorrelation and partial autocorrelation functions of Δq_t unambiguously suggests the use of an MA(2) specification (corresponding plots are available on request).

The ETS model class was developed by Hyndman et al. (2002, 2008) and encompasses various exponential smoothing methods within a theoretically founded state-space framework which is estimated recursively by employing maximum-likelihood methods. The ETS framework consists of a signal equation for the forecast variable and a number of state equations for the components that cannot be observed (level, trend, seasonal, Hyndman & Athanasopoulos, 2013, Section 7/7). Since the present data are given in logs and have been deseasonalized and stationarized, the appropriate ETS model is ETS(A, N, N) with **A**dditive error, **N**o trend component and **N**o seasonal component, thereby corresponding to classical single exponential smoothing. In state-space form, ETS(A, N, N) reads:

$$\Delta q_t = l_{t-1} + \upsilon_t, \tag{6}$$

$$l_t = l_{t-1} + \gamma \upsilon_t, \tag{7}$$

with Eq. (6) representing the signal equation, Eq. (7) the state equation, l_t denoting the unobservable level component, γ with $0 \leqslant \gamma \leqslant 1$ the smoothing parameter, and υ_t denoting an error term to be i.i.d. $\sim N(0, \sigma_\upsilon^2)$ appearing in both the signal and the state equation (Hyndman & Athanasopoulos, 2013, Section 7/7, so-called single-source-of-error model). The estimation of the system of Eqs. (6) and (7) is undertaken by employing maximum-likelihood methods.

Finally, the Naïve-1 or absolute no change model, which assumes that the forecast value of a variable in period t is equal to its realized value in period $t-1$ (one-step-ahead forecasting), serves as a natural benchmark forecast model:

$$\Delta\hat{q}_t^{naive} = \Delta q_{t-1}, \tag{8}$$

where the "hatted" variable on the left-hand side of Eq. (8) corresponds to the forecast value produced by this very model. Consequently, MA(2), ETS(A, N, N), and Naïve-1 models are employed as univariate benchmarks. This results in a total of seven rival forecast models whose forecast accuracy is evaluated below.

Forecast combination methods

Four forecast combination methods proposed in the literature are employed, whereby the combined forecast for tourist arrivals to Vienna $\hat{q}_t^c$ is defined as follows:

$$\Delta\hat{q}_t^c = \sum_{m=1}^{M} w^m \Delta\hat{q}_t^m, \tag{9}$$

with $\hat{q}_t^m$ denoting the individual forecast values produced by the $m = 1, \ldots, M$ rival forecast models and w^m the weights these individual forecast values are assigned concerning their contribution to the combined forecast. Forecast combination is not a new phenomenon in tourism demand forecasting. Usually, two approaches of how to specify the weights w^m in Eq. (9) are common across studies (e.g. Andrawis et al., 2011; Cang, 2011; Chan et al., 2010; Shen et al., 2011; Song et al., 2009; Wong et al., 2007; Wu & Zhang, 2014) besides other, usually more sophisticated, suggestions: simple uniform weights and some form of weights negatively related to the (square) forecast error in the spirit of Bates and Granger (1969).

Thus, for simple uniform weights, Eq. (9) simplifies to:

$$\Delta\hat{q}_t^c = \sum_{m=1}^{M} \frac{1}{M} \Delta\hat{q}_t^m, \tag{10}$$

with $w^m = 1/M$. Despite its simplicity, the simple uniform forecast combination as given by Eq. (10) has proven hard to beat by more sophisticated forecast combination methods in typical empirical forecast situations, i.e. when sample sizes are small (e.g. Clements & Hendry, 1998; Costantini & Kunst, 2011; Genre, Kenny, Meyler, & Timmermann, 2013; Timmermann, 2006).

For weights negatively related to the square forecast error, so-called Bates–Granger weights (Bates & Granger, 1969), Eq. (9) can be rewritten as:

$$\Delta\hat{q}_t^c = \sum_{m=1}^{M} \frac{(1/MSE_T^m)}{\left(\sum_{m=1}^{M} 1/MSE_T^m\right)} \Delta\hat{q}_t^m, \tag{11}$$

with $w^m = (1/MSE_T^m)/(\sum_{m=1}^{M} 1/MSE_T^m)$, while $1/MSE_T^m$ denotes the inverse of the mean square error of forecast model m observed over the forecast sample T. In the wider forecasting literature, this forecast combination method as described by Eq. (11), which assigns weights according to the observed track record in forecasting, has been shown to perform well in certain forecasting situations (e.g. de Menezes & Bunn, 1993; Hsiao & Wan, 2014).

Another possibility of observed track record based forecast combination, which has not drawn the attention of many researchers in tourism demand forecasting so far, is determining the weights w^m based on the multiple forecast encompassing F-test, which was introduced by Harvey and Newbold (2000). Only Shen et al. (2011) have employed such forecast encompassing tests for tourism demand forecast combination, albeit solely for forecast horizon $h = 1$.

In doing so, $m = 1, \ldots, M$ rival forecast models that produce out-of-sample forecast errors $e_t^m = \Delta q_t - \Delta\hat{q}_t^m$ for the forecast variable of interest Δq_t have to be considered. Then, the encompassing test procedure uses M encompassing regressions:

$$\begin{aligned}
e_t^1 &= a_1(e_t^1 - e_t^2) + a_2(e_t^1 - e_t^3) + \cdots + a_{M-1}(e_t^1 - e_t^M) + \eta_t^1, \\
e_t^2 &= a_1(e_t^2 - e_t^1) + a_2(e_t^2 - e_t^3) + \cdots + a_{M-1}(e_t^2 - e_t^M) + \eta_t^2, \\
&\vdots \\
e_t^M &= a_1(e_t^M - e_t^1) + a_2(e_t^M - e_t^2) + \ldots + a_{M-1}(e_t^M - e_t^{(M-1)}) + \eta_t^M.
\end{aligned} \tag{12}$$

These M homogeneous regressions (12) yield M regression F-statistics. A forecast model m is said to encompass the rival forecast models if the F-statistic in the regression with dependent variable e_t^m is insignificant. Following the evidence of the forecast encompassing tests, simple uniform forecast combinations are obtained according to the following rule. If F-tests reject or accept their null hypotheses in all M regressions, a new forecast will be formed as a uniformly weighted average of all model-based predictions: $w^m = 1/M$ as in Eq. (10). If some, say $m < M$, F-tests reject their null hypotheses, only those $M - m$ models that encompass their rival forecast models are combined. In this case, each of the surviving models receives a weight of $w^m = 1/(M - m)$ (Costantini et al., 2016).

Finally, the fourth and last forecast combination method proposed in this study is the application of Bates–Granger weights as given in Eq. (11) to the rival forecast models that have survived the forecast encompassing tests as described by Eqs. (12), thereby eliminating highly inaccurate forecast models in the first step to prevent them from having any detrimental impact on the combined forecast in Bates–Granger weighting in the second step. This combination of two track record based forecast combination methods has only been rarely employed in the forecasting literature (Costantini et al., 2016; Shen et al., 2011, are the only two known examples, while solely the first of these two studies examines the properties of this novel forecast combination method in a systematic way for both lab and real-world data and across forecast horizons), thus investigating its potential merits deserves researchers' attention.

Forecast evaluation

The forecast evaluation procedure is carried out as follows. The forecast performance of the seven single forecast models (the four VAR specifications and the three univariate benchmarks) as well as of the four forecast combination methods is evaluated in terms of the accuracy of *ex-ante* out-of-sample predictive means of tourist arrivals to Vienna for forecast horizons $h = 1, 2, 3, 6, 12$ months ahead while using expanding estimation windows (or recursive forecasting). This means all 11 variables are forecast jointly. Pesaran and Pick (2011) find that averaging forecasts over different estimation windows almost always leads to a lower RMSE relative to forecasts that are based on single estimation windows, even in the presence of structural breaks.

Each forecast model is re-estimated on a monthly basis starting with sub-sample $2009M8 - 2012M10$ (39 observations) up to $2009M8 - 2014M9$ (62 observations for $h = 1$), ..., $2009M8 - 2013M10$ (51 observations for $h = 12$), respectively. Altogether, this delivers $T_1 = 24$ (for $h = 1$), ..., $T_{12} = 13$ (for $h = 12$) counterfactual observations per model since the forecast window is set to the period $2012M11 - 2014M10$.

As measures of forecasting accuracy, the traditional RMSE and MAE are employed; the latter of which is more sensitive to small deviations from zero. The RMSE and MAE values per forecast model, forecast combination method, and forecast horizon including rankings from (1) (best performance) to (11) (worst performance) are given in parentheses in Table 3. The overall best-performing forecast models or forecast combination methods, respectively, are given in boldface, whereas the best performing VAR specifications are indicated by gray-shaded cells.

Moreover, the Hansen test on superior predictive accuracy is calculated to investigate if the Naïve-1 benchmark and the over-parameterized VAR(2) model are statistically significantly outperformed by at least one of the remaining six forecast models (Hansen, 2005). The null hypothesis of the Hansen test is that a benchmark forecast model is not outperformed by any other forecast model. The Hansen consistent p-values, which are also reported in Table 3, indicate a rejection of the null hypothesis and a statistically significant outperformance of the VAR(2) model and/or the Naïve-1 benchmark at least at the 10% level in case they are given in boldface. Because of the limited number of observations available, these tests have not been evaluated for the forecast horizon $h = 12$.

Table 3 shows the results of the forecast evaluation of the single forecast models as well as the forecast combination methods. As can be seen, the VAR(2) and the Naïve-1 models frequently perform worst, which implies that the over-parameterization of the classical VAR(2) estimated by OLS indeed has a negative impact on its forecast accuracy as presumed. Simply assuming that total tourist arrivals to Vienna tomorrow will correspond to tourist arrivals today would therefore be a bad suggestion. Moreover, Hansen test results indicate that the other forecast models used in this study significantly outperform these two forecast models on a regular basis.

For shorter forecast horizons ($h = 1, 2$), one univariate model, MA(2), performs best in terms of both the RMSE and the MAE. For longer horizons ($h = 3, 6, 12$), however, combined forecasts based on (a) Bates–Granger weights, (b) forecast encompassing tests with uniform weights, and (c) forecast encompassing tests with Bates–Granger weights produce the lowest RMSE and MAE values, while for $h = 12$ and MAE the single BVAR(2) model outperforms all its competitors.[3] Among the multivariate models only, the FAVAR(2) performs best in 6 cases, whereas the BVAR(2) and the BFAVAR(2) perform best in 2 cases each.

This underlines the predictive ability of the 10 Google Analytics website traffic indicators, which represent the foraged information from the website of the DMO of Vienna, and the viability of the employed big data shrinking techniques. The good performance of the more sophisticated multivariate forecast models and, in particular, of the more sophisticated forecast combination methods including this additional information but excluding the impact of bad single forecast models (in particular VAR(2)) is noteworthy since it is at odds with the traded wisdom that for forecast horizons up to two years univariate forecast models are expected to produce the most accurate forecasts (Frechtling, 2001).

[3] For $h = 1$: FAVAR(2) and MA(2) survived the forecast encompassing tests; for $h = 2$: BVAR(2), FAVAR(2), BFAVAR(2), MA(2), ETS(A, N, N), and Naïve-1; for $h = 3$: BVAR(2), FAVAR(2), BFAVAR(2), MA(2), and ETS(A, N, N); for $h = 6$: all forecast models; and for $h = 12$: BVAR(2), FAVAR(2), BFAVAR(2), MA(2), ETS (A, N, N), and Naïve-1 (detailed test results are available on request).

Table 3
RMSE, MAE values and Hansen test results.

		$h=1$ RMSE	MAE	$h=2$ RMSE	MAE	$h=3$ RMSE	MAE	$h=6$ RMSE	MAE	$h=12$ RMSE	MAE
Single forecast models	VAR(2)	0.03616 (9)	0.03002 (8)	0.05884 (11)	0.04619 (11)	0.06117 (11)	0.04978 (11)	0.05444 (9)	0.04689 (9)	0.09427 (11)	0.06746 (11)
	BVAR(2)	0.03375 (7)	0.02950 (7)	0.04232 (8)	0.03750 (7)	0.04449 (6)	0.04070 (9)	0.04489 (5)	0.04123 (7)	0.04724 (4)	**0.04267** (1)
	FAVAR(2)	0.02865 (4)	0.02328 (4)	0.04178 (6)	0.03631 (2)	0.04507 (7)	0.04052 (5)	0.04576 (7)	0.04120 (5)	0.04771 (8)	0.04288 (5)
	BFAVAR(2)	0.03424 (8)	0.03010 (9)	0.04224 (7)	0.03757 (8)	0.04433 (4)	0.04055 (6)	0.04488 (4)	0.04121 (6)	0.04724 (5)	0.04269 (3)
	MA(2)	**0.02554** (1)	**0.02084** (1)	**0.04093** (1)	**0.03627** (1)	0.04430 (2)	0.04037 (2)	0.04490 (6)	0.04120 (4)	0.04728 (7)	0.04269 (2)
	ETS(A, N, N)	0.04285 (10)	0.03786 (10)	0.04316 (9)	0.03851 (9)	0.04431 (3)	0.04039 (3)	0.04485 (3)	0.04120 (3)	0.04721 (3)	0.04271 (4)
	Naïve-1	0.07693 (11)	0.06581 (11)	0.04788 (10)	0.04015 (10)	0.05783 (10)	0.04520 (10)	0.05143 (8)	0.04207 (8)	0.05321 (10)	0.04667 (10)
Hansen consistent p-values	Relative to VAR(2)	**0.02190**		**0.09240**		**0.06320**		0.19070			
	Relative to Naïve-1	**0.00030**		0.20860		**0.06170**		0.21920			
Forecast combination methods	Uniform weights	0.03278 (6)	0.02872 (6)	0.04153 (5)	0.03660 (6)	0.04508 (8)	0.04060 (7)	0.04406 (2)	0.04017 (2)	0.04842 (9)	0.04456 (9)
	Bates-Granger weights	0.02940 (5)	0.02507 (5)	0.04146 (4)	0.03645 (4)	0.04529 (9)	**0.04025** (1)	**0.04399** (1)	**0.04002** (1)	0.04728 (6)	0.04392 (8)
	FE plus uniform weights	0.02658 (3)	0.02193 (3)	0.04103 (3)	0.03648 (5)	**0.04422** (1)	0.04060 (8)	N/A	N/A	0.04671 (2)	0.04314 (6)
	FE plus Bates-Granger weights	0.02640 (2)	0.02178 (2)	0.04100 (2)	0.03642 (3)	0.04436 (5)	0.04049 (4)	N/A	N/A	**0.04657** (1)	0.04327 (7)

Source: TourMIS, Google Analytics data for www.wien.info, and own calculations. Minimum RMSE and MAE values per forecast horizon across single forecast models and forecast combination methods are given in boldface, whereas the best performing VAR specifications are indicated by gray-shaded cells. "FE" denotes forecast combination based on forecast encompassing test results, while "N/A" indicates that none of the regression F-statistics were significant for $h = 6$. Boldface Hansen consistent p-values denote rejection of the null hypothesis of no outperformance of the VAR(2) and Naïve-1 benchmarks by at least one competing forecast model at the 10% level or higher. Square forecast losses are assumed to be minimized when calculating the Hansen test statistics.

Conclusion

Website traffic data such as Google Analytics website traffic indicators can be used to develop forecasts which enhance managerial decisions. In this study, 10 Google Analytics website traffic indicators (Average Session Duration, Average Time on Page, Bounce Rate, New Sessions, Page Views, Returning Visitors, Social Network Referrals, Total Sessions, Unique Page Views, and Users) were used to predict tourism demand in terms of total tourist arrivals to Vienna. Since the sample consisted of 11 variables, including total arrivals, big data shrinking methods became necessary, in particular with regards to the short time-series dimension (2008M08 to 2014M10) of the sample. The big data shrinkage methods geared towards the employed VAR model class were Bayesian estimation of the VAR (BVAR), extraction of common factors from the 10 Google Analytics website traffic indicators to reduce the dimensionality of the VAR (FAVAR), and a recent fusion of Bayesian and factor analysis (BFAVAR). The last of these was here applied to tourism demand forecasting for the first time and, together with BVAR and FAVAR, proved quite accurate when evaluating its forecast performance against the classical VAR.

While one univariate benchmark model, MA(2), outperformed all other models for shorter forecast horizons of $h = 1, 2$ months ahead, the other two univariate benchmarks, ETS and Naïve-1, were not that convincing. In fact, the latter benchmark was often significantly outperformed by its competitors across forecast horizons, as was the classical VAR. For longer forecast horizons of $h = 3, 6, 12$ months ahead, excluding $h = 12$ where BVAR performed best according to MAE, forecast combination methods such as Bates–Granger weights, forecast encompassing tests with uniform weights, and forecast encompassing tests with Bates–Granger weights outperformed the single forecast models as well as combined forecasts based on simple uniform weights. The combination of forecast encompassing tests with Bates–Granger weights was used in tourism demand forecasting for the first time.

The results showed that forecast combination methods outperformed the univariate benchmark models for longer forecast horizons, which are typically more difficult to predict. In addition, apart from almost immediate forecast horizons, forecast accuracy of tourism demand already benefits from the foraged information contained in the Google Analytics website traffic indicators for forecast horizons shorter than two years.

The use of big data in forecasting tourism demand is a new and important step for tourism research and the tourism industry. Since Google Analytics website traffic indicators were found to be useful in forecasting and the data are available as of today, nowcasting represents a new opportunity for the tourism industry. This study indicates that Google Analytics, or any other type of website traffic data, can not only be used by the IT department of DMOs such as the Vienna Tourist Board to enhance website performance, but also by forecasters to predict actual tourism demand to a destination. As a result of improved forecasting accuracy, tourism resources can be more efficiently allocated, cost reductions can be realized, and trends in tourist behavior can be detected earlier.

The main limitations of this research are as follows. Due to the private nature of Google Analytics data the study focuses on Vienna only. Thus, the results may differ for other European city destinations such as London, Madrid, Paris, or Rome. Nonetheless, the econometric methodology used in this study can also be applied to those cities, provided that their DMOs also use Google Analytics to track their websites' traffic. In principle, Google Analytics offers more website traffic indicators than those that were used in this study, in particular data on user demographics, such from which countries the users accessed the DMO website. Unfortunately, the authors of this study did not have access to this information.

Thus, future research could focus on distinguishing the domestic and the most important foreign source markets of a destination to determine appropriate forecast models at the source market level. This information would enable the tourism demand of specific source markets to be better analyzed and forecast, and inform the design and launch of customized (online) marketing campaigns. Also, since the forecast accuracy of models characterized by the use of Google Analytics indicators and big data shrinkage increases relative to the univariate benchmarks for the longer forecast horizons, once enough data become available, future research could include forecast horizons of more than one year as well. In the future, it may be worthwhile to investigate the impact of different calibrations of the Minnesota prior and/or the impact of the use of other popular informative priors such as the Normal-Wishart prior, the Sims-Zha normal-Wishart prior, or the Sims-Zha normal-flat prior (see e.g. EViews, 2013; Karlsson, 2013, for overviews) on the forecast performance of the B(FA) VAR models as well.

Acknowledgement

The authors would like to thank the three reviewers of this journal, the participants of the 9th International Conference on Computational and Financial Econometrics (London, UK) and of the 35th Annual International Symposium on Forecasting (Riverside, US) for helpful comments and suggestions for improvement. In addition, they would like to express their special gratitude to Clemens Költringer and the Vienna Tourist Board for granting access to the data.

References

Amir-Ahmadi, P., & Uhlig, H. (2013). Measuring the Effects of a Shock to Monetary Policy: A Bayesian Factor-Augmented VAR Approach with Sign Restrictions. Mimeo.

Andrawis, R. R., Atiya, A. F., & El-Shishiny, H. (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting, 27*, 870–886.

Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting, 27*, 822–844.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian Vector Auto Regressions. *Journal of Applied Econometrics, 25*, 71–92.

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.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly, 20*, 451–468.

Beaulieu, J. J., & Miron, J. A. (1993). Seasonal unit roots in aggregate U.S. data. *Journal of Econometrics, 55*, 305–328.

Bekiros, S. D., & Paccagnini, A. (2014). Bayesian forecasting with small and medium scale factor-augmented vector autoregressive DSGE models. *Computational Statistics and Data Analysis, 71*, 298–323.

Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics, 120*, 387–422.

Boswell, P. (2011). Google Analytics: Measuring Content Use and Engagement. Paper presented at the Technical Communication Summit, Sacramento, California.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis, forecasting and control*. San Francisco: Holden Day.

Breña Moral, J.A. (2007). Use Google Analytics and R to Predict Web Traffic. URL: http://www.juanantonio.info/docs/2007/googleAnalyticsAndRpart1.pdf. Accessed on 06/14/2016.

Cang, S. (2011). A non-linear tourism demand forecast combination model. *Tourism Economics, 17*, 5–20.

Chan, K. C., Witt, S. F., Lee, Y. C. E., & Song, H. (2010). Tourism forecast combination using CUSUM technique. *Tourism Management, 31*, 891–897.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*, 1165–1188.

Chow, H.K., & Choy, K.M. (2009). Monetary Policy and Asset Prices in a Small Open Economy: A Factor-Augmented VAR Analysis for Singapore. SMU Working Paper No. 11–2009. Singapore: Singapore Management University (SMU).

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*, 559–583.

Clements, M., & Hendry, D. F. (1998). *Forecasting economic time series*. Cambridge: Cambridge University Press.

Costantini, M., Gunter, U., & Kunst, R. M. (2016). Forecast Combinations in a DSGE-VAR Lab. *Journal of Forecasting*. http://dx.doi.org/10.1002/for.2427.

Costantini, M., & Kunst, R. M. (2011). Combining forecasts based on multiple encompassing tests in a macroeconomic core system. *Journal of Forecasting, 30*, 579–596.

de Menezes, L., & Bunn, D. W. (1993). Diagnostic tracking and model specification in combined forecast of U.K. *Inflation. Journal of Forecasting, 12*, 559–572.

Doan, T. A., Litterman, R. B., & Sims, C. A. (1984). Forecasting and conditional projections using realistic prior distributions. *Econometric Reviews, 1*, 1–100.

European Cities Marketing (ECM) & MODUL University Vienna (MU) (2014). The European Cities Marketing Benchmarking Report, 10th Official Edition 2013–2014.

EViews (2013). *EViews 8 Users Guide II*. Irvine, CA: IHS Global Inc.

Frechtling, D. C. (2001). *Forecasting tourism demand: Methods and strategies*. Burlington, MA: Butterworth Heinemann.

Gawlik, E., Kabaria, H., & Kaur, S. (2011). Predicting tourism trends with Google Insights. URL: http://cs229.stanford.edu/proj2011/GawlikKaurKabaria-PredictingTourismTrendsWithGoogleInsights.pdf. Accessed on 06/14/2016.

Genre, V., Kenny, G., Meyler, A., & Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting, 29*, 108–112.

Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *The Review of Economics and Statistics, 97*, 436–451.

Google (2015). Bounce rate. URL: https://support.google.com/analytics/answer/1009409?hl=en&vid=0-635755843569085164-14125533301200080186. Accessed on 08/19/2015.

Granowitz, A. (2014). Using Google Analytics with R. URL: https://developers.google.com/analytics/solutions/r-google-analytics. Accessed on 06/14/2016.

Gunter, U., & Önder, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data. *Tourism Management, 46*, 123–135.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics, 23*, 365–380.

Harvey, D. I., & Newbold, P. (2000). Tests for multiple forecast encompassing. *Journal of Applied Econometrics, 15*, 471–482.

Hsiao, C., & Wan, S. K. (2014). Is there an optimal forecast combination? *Journal of Econometrics, 178*, 294–309.

Huang, X., Zhang, L., & Ding, Y. (2016). The Baidu Index: Uses in predicting tourism flows – A case study of the Forbidden City. *Tourism Management*. http://dx.doi.org/10.1016/j.tourman.2016.03.015.

Hylleberg, S., Engle, R. F., Granger, C. W. J., & Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of Econometrics, 44*, 215–238.

Hyndman, R.J., & Athanasopoulos, G. (2013). Forecasting: principles and practice, Section 7/7. URL: http://otexts.com/fpp/7/7. Accessed on 10/13/2015.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Berlin, Heidelberg: Springer.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting, 18*, 439–454.

Jackman, M., & Natiram, S. (2015). Research note: Nowcasting tourist arrivals in Barbados – Just Google it! *Tourism Economics, 21*, 1309–1313.

Jansen, B. J., Ciamacca, C. C., & Spink, A. (2008). An analysis of travel information searching on the web. *Information Technology and Tourism, 10*, 1–18.

Karlsson, S. (2013). Forecasting with Bayesian vector autoregression. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting, Part B* (2, pp. 791–897). Oxford: Elsevier.

Kent, M. L., Carr, B. J., Husted, R. A., & Pop, R. A. (2011). Learning web analytics: A tool for strategic communication. *Public Relations Review, 37*, 536–543.

Kulendran, N., & King, M. L. (1997). Forecasting international quarterly tourism flows using error correction and time series models. *International Journal of Forecasting, 13*, 319–327.

Kulendran, N., & Witt, S.F. (1997). Modelling UK Outbound Tourism Demand. Paper presented at the 17th International Symposium on Forecasting. Bridgetown, Barbados.

Kulendran, N., & Witt, S. F. (2001). Cointegration versus least squares regression. *Annals of Tourism Research, 28*, 291–311.

Li, G., Song, H., & Witt, S. F. (2006). Forecasting tourism demand using econometric models. In D. Buhalis & C. Costa (Eds.), *Tourism Management Dynamics: Trends, Management and Tools* (pp. 219–228). Oxford: Elsevier.

Lim, C., & McAleer, M. (2001). Cointegration analysis of quarterly tourism demand by Hong Kong and Singapore for Australia. *Applied Economics, 33*, 1599–1619.

Litterman, R. B. (1986). A statistical approach to economic forecasting. *Journal of Business and Economic Statistics, 4*, 1–4.

Ljubaj, I. (2012). Estimating the Impact of Monetary Policy on Household and Corporate Loans: a FAVEC Approach. CNB Working Paper No. W-34. Zagreb: Croatian National Bank (CNB).

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin, Heidelberg: Springer.

Meares, C. (2015). What is Online Engagement And Why Should You Measure It? (Part One). URL: http://www.maassmedia.com/blog/what-is-online-engagement-and-why-should-you-measure-it-part-one/. Accessed on 08/19/2015.

New Zealand Institute of Economic Research (NZIER) (2012). Integrating big data to forecast New Zealand's visitor arrivals: a structured approach. NZIER – Tourism forecasts 2012–2018 – technical report.

Oh, C. O. (2005). The contribution of tourism development to economic growth in the Korean economy. *Tourism Management, 26*, 39–44.

Önder, I., & Gunter, U. (2016). Forecasting tourism demand with google trends for a major european city destination. *Tourism Analysis, 21*, 203–220.

Pakkala, H., Presser, K., & Christensen, T. (2012). Using Google Analytics to measure visitor statistics: The case of food composition websites. *International Journal of Information Management, 32*, 504–512.

Pang, I.A.J. (2010). Forecasting Hong Kong economy using factor augmented vector autoregression. MPRA Paper No. 32495. Munich: Munich Personal RePEc Archive (MPRA).

Pesaran, M. H., & Pick, A. (2011). Forecast combination across estimation windows. *Journal of Business and Economic Statistics, 29*, 307–318.

Pirolli, C., & Card, S. (1999). Information foraging theory. *Psychological Review, 106*, 643–675.

Plaza, B. (2011). Google Analytics for measuring website performance. *Tourism Management, 32*, 477–481.
Ribon, S. (2011). The Effect of Monetary Policy on Inflation: A Factor Augmented VAR Approach using disaggregated data. BOI Discussion Paper No. 2011.12. Jerusalem, Tel Aviv: Bank of Israel (BOI).
Shan, J., & Wilson, K. (2001). Causality between trade and tourism: Empirical evidence from China. *Applied Economics Letters, 8*, 279–283.
Shen, S., Li, G., & Song, H. (2011). Combination forecasts of international tourism demand. *Annals of Tourism Research, 38*, 72–89.
Sims, C. (1980). Macroeconomics and reality. *Econometrica, 48*, 1–48.
Song, H., Romilly, P., & Liu, X. (2000). An empirical study of outbound tourism demand in the UK. *Applied Economics, 32*, 611–624.
Song, H., Smeral, E., Li, G., & Chen, J. L. (2013). Tourism forecasting using econometric models. In C. Costa, E. Panyik, & D. Buhalis (Eds.), *Trends in European Tourism Planning and Organisation* (pp. 289–309). Bristol: Channel View.
Song, H., & Witt, S. F. (2006). Forecasting international tourist flows to Macau. *Tourism Management, 27*, 214–224.
Song, H., Witt, S. F., & Jensen, T. C. (2003). Tourism forecasting accuracy of alternative econometric models. *International Journal of Forecasting, 19*, 123–141.
Song, H., Witt, S. F., & Li, G. (2009). *The advanced econometrics of tourism demand*. New York, London: Routledge.
Song, H., Witt, S. F., Wong, K. F., & Wu, C. (2009). An empirical study of tourism demand forecast combination. *Journal of Hospitality and Tourism Research, 33*, 3–29.
Stock J.H., & Watson, M.W. (1998). Diffusion indexes. NBER Working paper No. 6702. Cambridge, MA: National Bureau of Economic Research (NBER).
Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics, 20*, 147–162.
Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (pp. 135–196). Amsterdam: Elsevier.
Vadera, P. (2013). Forecasting the number of visitors on your website using R. URL: http://www.tatvic.com/blog/forecasting-the-number-of-visitors-on-your-website-using-r-part-i/. Accessed on 06/14/2016.
W3Techs (2015). Usage of traffic analysis tools for websites. URL: http://w3techs.com/technologies/overview/trafficanalysis/ all Usage of traffic analysis tools for websites. Accessed on 10/15/2015.
Witt, S. F., Song, H., & Louvieris, P. (2003). Statistical testing in forecasting model selection. *Journal of Travel Research, 42*, 151–158.
Wong, K. K. F., Song, H., & Chon, K. S. (2006). Bayesian models for tourism demand forecasting. *Tourism Management, 27*, 773–780.
Wong, K. K. F., Song, H., Witt, S. F., & Wu, D. C. (2007). Tourism forecasting: To combine or not to combine? *Tourism Management, 28*, 1068–1078.
Wu, L., & Zhang, J. (2014). The variance-covariance method using IOWGA operator for tourism forecast combination. *International Journal of Supply and Operations Management, 1*, 152–166.
Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management, 46*, 386–387.
Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research, 53*, 433–447.
Zwitch, A. (2013). Using R and the Google Analytics API. URL: http://randyzwitch.com/r-google-analytics-api/. Accessed on 06/14/2016