Journal of Travel Research
http://jtr.sagepub.com/

Forecasting Short Time-Series Tourism Demand with Artificial Intelligence Models

Gongmei Yu and Zvi Schwartz

Journal of Travel Research 2006 45: 194
DOI: 10.1177/0047287506291594

The online version of this article can be found at:
http://jtr.sagepub.com/content/45/2/194

Published by:

http://www.sagepublications.com

On behalf of:

Travel and Tourism Research Association

Additional services and information for *Journal of Travel Research* can be found at:

Email Alerts: http://jtr.sagepub.com/cgi/alerts

Subscriptions: http://jtr.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://jtr.sagepub.com/content/45/2/194.refs.html

Downloaded from jtr.sagepub.com at Universita della Calabria on March 8, 2011

Forecasting Short Time-Series Tourism Demand with Artificial Intelligence Models

GONGMEI YU AND ZVI SCHWARTZ

This study examines the forecast accuracy of fuzzy time series and grey theory in predicting annual U.S. tourist arrivals. The performance of the two artificial intelligence (AI) models is compared to that of two simple methods—double moving average and double exponential smoothing. The rigorous testing approach includes a large sample stratified to adequately represent four generic trend patterns: a rolling short-term forecast, a large holdout sample, models fitting with both equal number of years and optimal number of years, and tests of statistical significance using Wilcoxon's signed-ranks nonparametric test. This study's findings indicate, in contrast to recent findings, that the complicated models are not likely to generate a more accurate forecast than the simple traditional models. Given the notable cost associated with these AI forecasting methods, our finding of no significant accuracy advantage suggests that tourism forecasters should not rush to adopt these two methods without careful consideration.

Keywords: *demand forecasting; fuzzy time series; grey theory; annual tourist arrivals*

Tourism demand forecasting has gained increased attention in the past 3 decades, and the accuracy of time-series and econometric models became the focus of extensive research efforts as providers and policy makers recognized the industry's size, growth rate, and economic impact (Witt and Witt 1992). The need for more accurate forecasts of tourism demand is driven by the desire to reduce risk and uncertainty. This need, according to Frechtling (2001), is further accentuated by the perishable nature of tourism products, the simultaneous purchase-production process, and the role of complementary services in shaping consumer satisfaction and, thereby, their subsequent demand for future tourism services.

Although research efforts address mostly conventional tourism-forecasting methods (e.g., Song and Witt 2003; Lim and McAleer 2001; Schwartz and Cohen 2004a, 2004b; Witt, Song, and Louvieris 2003; Wong 2002), in recent years there has been an emerging interest in artificial intelligence (AI) forecasting models. Research indicates that these AI forecasting methods—including neural networks, rough sets theory, fuzzy time-series theory, grey theory, genetic algorithms, and expert systems—sometimes outperform conventional forecasting methods (Wang 2004). Applications of AI methods in tourism demand forecasting include, for example, Law and Au (1999) and Law (2000), who used a feed-forward neural network and a back-propagation neural network to model the complex relationship between tourism demand and its determinants. These studies found that neural networks were more accurate than regression and time-series models. Rough sets theory was applied by Au and Law (2000) to forecast sightseeing expenditures in Hong Kong. Goh and Law (2003) derived decision rules from an information table of tourist arrivals with a rough sets approach, in which the induced rules forecasted changes in demand with 87% accuracy.

This study focuses on *fuzzy time-series* and *grey theory* models. Both are designed to handle forecasts when few past observations are available. *Fuzzy time series* was first proposed by Song and Chissom (1993, 1994), who used first-order time-invariant and first-order time-variant fuzzy time-series models to forecast enrollment at the University of Alabama. Hwang, Chen, and Lee (1998) simplified the arithmetic operation process and found that the average forecasting error is smaller than that of the Song and Chissom model. *Grey theory* models have been used successfully in various prediction applications such as integrated circuits and the air travel industry (Lin, Su, and Hsu 2001; Hsu 2003). In a recent article, Wang (2004) demonstrated the use of grey theory and fuzzy time series in forecasting tourist arrivals to Taiwan from Hong Kong, the United States, and Germany.

Indeed, tourism data often comprise short time series, and, as such, the two techniques seem most adequate for predicting tourism demand. Both fuzzy logic and grey theory are, however, far from easy to implement. They are considerably more complicated than traditional forecasting methods. In particular, fuzzy forecasting requires that the numerical tourism data be transformed into a linguistic scale and then back into numbers. Another obstacle for widespread application is that the two methods are not yet available in commercial statistical or forecasting packages, thus requiring that the researcher or forecaster write the computer code. Hence, the purpose of this study is to further explore the appropriateness of fuzzy time series and grey theory as tourism demand forecasting methods. Specifically, it attempts to determine whether these complicated methods are indeed more accurate than

Gongmei Yu is a graduate student in the Department of Recreation, Sport, and Tourism at the University of Illinois at Urbana-Champaign. Zvi Schwartz, PhD, is an assistant professor in the Department of Recreation, Sport, and Tourism at the University of Illinois at Urbana-Champaign.

Journal of Travel Research, Vol. 45, November 2006, 194-203
DOI: 10.1177/0047287506291594
© 2006 Sage Publications

Downloaded from jtr.sagepub.com at Universita della Calabria on March 8, 2011

TABLE 1
ANNUAL TOURIST ARRIVALS FROM EIGHT COUNTRIES TO THE UNITED STATES

	Country							
Year	Canada	Germany	Brazil	Luxembourg	United Kingdom	Denmark	Nigeria	Yugoslavia
1985	10,721,000	509,131	191,783	4,161	860,837	56,308	42,003	18,926
1986	10,793,000	669,845	257,934	5,362	1,133,683	68,519	46,284	22,623
1987	12,253,000	952,119	238,073	7,818	1,362,479	96,440	26,161	26,247
1988	13,700,000	1,153,359	298,738	9,681	1,827,831	100,482	26,808	28,377
1989	15,325,000	1,076,385	333,849	8,752	2,221,871	107,062	22,449	28,092
1990	17,263,000	1,202,826	398,484	10,532	2,243,792	96,796	19,444	29,203
1991	19,113,000	1,430,193	459,384	11,910	2,495,354	95,399	22,268	27,637
1992	18,598,461	1,691,663	475,266	13,738	2,823,983	100,353	19,001	15,395
1993	17,293,000	1,826,757	555,102	15,796	2,999,301	103,665	19,262	13,034
1994	14,974,000	1,622,835	652,412	15,245	2,772,345	101,062	16,082	19,508
1995	14,662,000	1,847,973	838,365	18,077	2,888,195	108,003	15,700	19,017
1996	15,301,000	1,996,824	848,453	19,192	3,246,237	137,307	19,256	11,089
1997	15,127,000	1,994,296	940,698	20,797	3,720,979	144,605	24,414	11,570
1998	13,422,000	1,901,938	909,477	19,706	3,974,976	138,689	28,010	10,149
1999	14,110,000	1,984,627	665,013	19,368	4,252,160	146,479	29,734	5,839
2000	14,594,000	1,786,045	737,245	16,385	4,703,008	149,211	27,124	8,039
2001	13,507,000	1,313,756	551,406	12,093	4,097,258	126,345	37,406	10,206
2002	12,968,103	1,189,856	405,094	9,646	3,816,736	118,716	40,128	10,972

Source: Office of Travel and Tourism Industries (1996, 2004).

traditional, simpler approaches when few data are available for model fitting. Assuming a difference exists, the study also sets out to examine whether that difference warrants the high cost (in skill and effort) associated with the application of the two AI methods.

A major contribution of this article is that it takes a more rigorous approach to testing the performance of the more complicated methods as well as the benchmark models. The number of tested series is large, the sample is stratified to adequately represent four generic trend patterns identified in the sample framework, the holdout sample and the rolling short-term forecast procedure attempt to better reflect the reality of forecasting annual tourists' arrivals, and, finally, the differences in the models' accuracy are tested for statistical significance. In addition, the benchmarks were selected in accordance with the principle of the best available extrapolative univariate procedure (Mincer and Zarnowitz 1969).

The article proceeds as follows: the first section discusses the data—annual tourist arrivals to the United States from various countries of origin. It portrays four generic patterns in tourism demand data and explains the rationale behind the selection of the eight time series (i.e., eight countries of origin) used for testing in this study. The second section outlines the way in which the forecast accuracy was measured, including magnitude and directional change error measures as well as the nonparametric Wilcoxon's signed-rank test. The third section describes the two tested methods, fuzzy logic and grey theory. The results section starts by examining the relation between parameters' values and the forecast accuracy. It then compares the accuracy of the tested models (fuzzy and grey) to that of the simple, more traditional benchmark forecasting models (double moving average and double exponential smoothing) and tests the differences in accuracy for statistical significance. Finally, the last section provides a summary, a discussion of the implications, the article's limitations, and possible avenues for future research.

DATA

The models' accuracy was tested using annual tourist arrivals to the United States from various countries from 1985 to 2002. The data are posted online as part of ITA, the Web site of the Office of Travel and Tourism Industries (2004). Examination of the arrival data, that is, the scatter plots of 87 different origin countries, reveals that more than 90% of the time-series charts exhibit some kind of trend pattern. The four following major patterns were identified:

An increase followed by a decrease (I/D): 56% of the countries
An increase (I): 28% of the countries
A decrease followed by an increase (D/I): 5% of the countries
A decrease (D): 3% of the countries

Together, the I/D and I trend patterns account for more than 84% of the time series in the data set of annual tourist arrivals to the United States. The D/I and D trend patterns are not as common, accounting for only 8% of the time series. The convenient sample of eight countries is stratified to reflect both the frequency of the four trend patterns in the sample frame of 87 countries and the variety of tourists' volume. Hence, the sample is as follows: four origin countries are characterized as I/D (Canada, Germany, Brazil, and Luxembourg) with varying tourist volume. Two countries are characterized as I trend pattern, of which the first (United Kingdom) exhibits a steady increase, whereas the second (Denmark) exhibits a steplike increase pattern. Arrivals of tourists from Nigeria represent the D/I trend pattern, and arrivals from Yugoslavia represent the D trend pattern. Table 1 displays the arrival figures for each of the eight origin countries included in the sample.

Downloaded from jtr.sagepub.com at Universita della Calabria on March 8, 2011

FORECAST ACCURACY EVALUATION

Out-of-Sample Tests of Forecast Accuracy

This study assesses the accuracy of the forecasting methods using out-of-sample tests rather than using goodness-of-fit to past data (Martin and Witt 1989; Tashman 2000). In the study, a process of model refitting is applied to better reflect the way in which annual tourism demand forecasting is performed in reality. Starting with a fitting series of M observations, the study continues by fitting to $M + 1$ observations, $M + 2$ observations, and so forth.

Forecasting Accuracy Measures

The forecast accuracy is measured with two commonly used error magnitude measures: mean absolute percentage error (MAPE) and root mean squared percentage error (RMSPE) (Martin and Witt 1989; Song and Witt 2000; Song, Witt, and Jensen 2003; Turner and Witt 2001). In addition, the Henriksson-Merton (HM) test of forecast rationality and positive value is applied to assess the directional change of the forecasts (Henriksson and Merton 1981; Witt, Song, and Louvieris 2003).

Test for Equal Forecasting Accuracy

The nonparametric Wilcoxon's matched-pairs signed-ranked test (Daniel 1990, pp. 150-55; Schwartz and Hiemstra 1997) is used to compare the accuracy of the forecasting methods. It is an alternative to the t-test in situations in which the assumption of a normal distribution is violated (such as this study's small sample size). The null hypothesis is that the loss differential has a median value of zero. If the test statistic $T(+)$ is smaller than the critical tabled value, one rejects the null hypothesis of equal forecasting accuracy.

FORECASTING METHODS

Although many robust forecasting approaches, such as autoregressive integrated moving average (ARIMA) and neural networks, require a large amount of historical data to train or fit the model, annual tourism demand data are often limited, restricting the choice of forecast methods. Grey theory and fuzzy time-series models, the focus of this investigation, have been suggested as adequate for short-term, current, and limited data; the grey theory model, for instance, can be fitted with as few as four observations (Chiang et al. 1998). The following section describes the two models and the relation between their fitted parameters and their accuracy. Both models were programmed using C. The computer code written for this study is available from the authors on request.

Fuzzy Time Series

The fuzzy time-series model, designed to forecast processes with linguistic value observations (Song and Chissom 1993), deals with the first-order difference of the time series, that is, the variation between this year (t) and the preceding one (t-1).

According to Hwang, Chen, and Lee (1998), the main assumption of the fuzzy time-series model is that the variation of this year is related to the variation of previous years. This relation is described by a set of heuristic rules (Hwang, Chen, and Lee 1998), and consequently the forecasting process is more complicated than traditional methods in which the relation is most often described by a function. When applied to tourism demand forecasting, the numerical value needs to be translated into a linguistic value, and then the predicted variation value (in linguistic terms) is translated back into a numerical value (Wang 2004).

The fuzzy time-series model's fundamental assumption is that the variation of the current year follows the trend of recent years and especially that of the previous year. In addition, the predicted variation is within a certain range that is determined by the historical variations. Consequently, if the actual variation is considerably different from recent trends or recent ranges, then the forecast error is likely to be large. Building a fuzzy time-series model involves human judgment when a decision about two parameters needs to be made. The forecaster first needs to select m, that is, the number of grades describing the extent of variation. For example, if m is 6, then the six variation grades could be *big decrease*, *decrease, no change*, *increase*, *big increase*, and *too big increase*. The second parameter is w, the window basis size denoting the number of previous observations used to generate the prediction.

Grey Theory

Grey theory, proposed by Deng (1982), is a generic theory that deals with systems that have poor, incomplete, and uncertain information. Unlike traditional time-series models that use the original data or the first-order differences, the grey model reduces randomness by using the accumulated generation operation (AGO) form. An exponential function is fitted based on a differential equation to estimate the future trend. A major advantage of the grey model is that it can be constructed based on a very short time series (as few as four items) and that no assumption about the statistical distribution of the data is necessary. The method's major disadvantage is that the data must be taken at equal intervals and in consecutive order (Lin, Su, and Hsu 2001).

THE BENCHMARK METHODS

The two benchmark models, *double moving average* and *double exponential smoothing*, were selected in accordance with the principle of the best available, simple extrapolative, univariate procedure. Moving average and exponential smoothing are commonly used techniques in business and economic forecasting (Box 1991; Snyder, Koehler, and Ord 2002). They are considered to be robust tourism-forecasting methods (Weatherford and Kimes 2003) and are often applied in tourism and hospitality demand forecasting. In addition, the two methods share two characteristics with fuzzy time series and grey theory: they are generally superior in short-term forecasting situations, and they often perform poorly when used with more random and nonlinear center-symmetry curve data (Wheelwright and Makridakis 1980; Lin, Su, and Hsu 2001). Advanced versions of the two

Downloaded from jtr.sagepub.com at Universita della Calabria on March 8, 2011

basic methods, namely, double moving average and double exponential smoothing, were adopted as the benchmarks because all tested time series show a trend element.

RESULTS

Parameters Value and the Forecast Accuracy

The selection of suitable parameters used in the model might affect the forecast accuracy. As noted by Hwang, Chen, and Lee (1998), however, it is not always easy to identify the relationship between the parameters and the forecast's average error. This section discusses the relationships identified in this study and what they imply about parameter values that are likely to generate more accurate forecasts.

Fuzzy Time Series

This study varies the window size, *w*, between 2 and 5, and the variation grade, *m*, between 4 and 9. The corresponding error magnitude (MAPE) and the probability of directionally correct forecasts ($\hat{p}_1 + \hat{p}_2$) are outlined in tables 2 and 3.

Although *m* seems to affect the accuracy, no clear pattern emerges. The accuracy of several countries is nonlinear in respect to *m*, rather than the expected linear relationship. Furthermore, the optimal value of *m* (i.e., the one that generates the smallest forecast errors) changes across countries. For example, when *m* equals 8, the MAPE and the RMSPE of Canada is the smallest. The same model has the most accurate forecast in Brazil when *m* equals 7. In addition, the curve may have two minima—in Nigeria, when *m* is equal to 5 and 9, the prediction is more accurate than with other values of *m*. The value of *m* for which the model generates the most accurate forecast may also change when using a different error measure. For Luxembourg, the best accuracy in terms of MAPE occurs when $m = 5$, whereas with RMSPE, $m = 6$ generates the most accurate forecast. Then again, many of the eight sampled countries show an almost identical pattern of relationship between *m* and the prediction's accuracy.

The impact of the window size, *w*, on the forecast accuracy is remarkably different than that of the grade size of variation, *m*. For some countries, such as the United Kingdom and Denmark, the difference in accuracy is rather small when *w* is varied. For countries that display an obvious accuracy change when *w* changes, the relationship is appropriately linear—the smaller the *w*, the smaller the MAPE and the RMSPE. Overall, the model generates the most accurate forecast when *w* equals 2.

The performance of the fuzzy time series in predicting directional change is not ideal (table 3). In many cases, $\hat{p}_1 + \hat{p}_2$ is less than 1. In other cases when $\hat{p}_1 + \hat{p}_2$ is greater than or equal to 1, $\alpha > 0.05$, that is, the forecasts are not rational. Among the few cases that $\hat{p}_1 + \hat{p}_2 \geq 1$ and the rationality test is significant, very few pass the positive value test ($\beta > 0.95$).

The impact of variation grade *m* and window size *w* on the direction change forecasting accuracy varies with different countries. For countries such as Luxembourg, Nigeria, and Yugoslavia, $\hat{p}_1 + \hat{p}_2$ changes slightly when *m* or *w* changes. The impact of variation grade on the $\hat{p}_1 + \hat{p}_2$ measure is larger than that of the window size for Denmark and the United Kingdom. The value of $\hat{p}_1 + \hat{p}_2$ changes considerably with *w* or *m* for other countries: for example, Canada's $\hat{p}_1 + \hat{p}_2$ ranges from 0.46 to 1.57. Overall, however, the relation between $\hat{p}_1 + \hat{p}_2$ and *m* or *w* is not clear.

To summarize, no clear pattern emerges to describe the relation between both the error magnitude and directional change accuracy and value of the parameters *m* and *w*. Furthermore, the choice of error measure (MAPE or RMSPE) does not seem to affect the choice of optimal parameter values.

Grey Theory

Although grey theory models can be fitted with as little as 4 data points, this study tested the model's one-step-ahead forecast accuracy on various lengths of time series, ranging from 4 to 7. The results are outlined in table 4. MAPE scores seem less sensitive to changes in *n* than do RMSPE scores. For countries such as Canada, Germany, and Denmark, the change in MAPE is rather small—less than 1% between the largest and the smallest values. The RMSPE scores, however, fluctuate considerably, with different *n* for all of the eight countries. Two relationship patterns between the forecasting accuracy and *n* were identified: a linear pattern, in which the larger the *n* value, the larger the MAPE or RMSPE; and a nonlinear pattern in which the most accurate forecast is generated when *n* equals 5 or 6 (Brazil and Yugoslavia). Note that the pattern of the relationship between the accuracy and *n* is not affected by the error measure (i.e., MAPE and RMSPE).

The directional change accuracy varies across countries. For countries such as Brazil and Luxembourg, all of the $\hat{p}_1 + \hat{p}_2$ values are greater than or equal to 1.0, and the forecasts are rational at $P < 0.05$. In most cases, however, the hypothesis about positive value is rejected at the 0.05 level. For other countries, many of the $\hat{p}_1 + \hat{p}_2$ terms are small; the rationality test is statistically significant in just a few cases, and only once is β larger than 0.95, meaning that only one forecast of directional change had a positive value.

The length of time series used to fit the model has a certain effect on the directional change forecasting accuracy. For countries such as Canada, Germany, and Brazil, the relation between $\hat{p}_1 + \hat{p}_2$ and *n* tends to be linear, but the sign of the relation is not consistent. For example, Canada's $\hat{p}_1 + \hat{p}_2$ decreases with *n*, whereas Germany's $\hat{p}_1 + \hat{p}_2$ increases with *n*. For other countries, the relation patterns are nonlinear.

In all, the forecasting accuracy (both the magnitude and the directional change errors) of the two AI models seems to depend, at least to some extent, on the size of the data set used to fit the model; however, no clear and consistent pattern emerged.

Comparisons of Forecasting Accuracy

The accuracy of the AI models is compared to that of the benchmarks using two approaches. The first compares the forecast accuracy with all models tested using the same data series length, *N*. That is, an equal number of historical observations was used across all models to fit each of the compared models. The second approach allows the compared models to be fitted on a series with different lengths. Each model is represented in the competition by the optimal or most accurate fitted version as identified in the previous sections. That is, the most accurate forecasts of each model (across all time-series length tested) are compared to the best performance of the other models. Hence, with the second approach, although

Downloaded from jtr.sagepub.com at Universita della Calabria on March 8, 2011

Downloaded from jtr.sagepub.com at Universita della Calabria on March 8, 2011

TABLE 2

FUZZY TIME-SERIES FORECAST ACCURACY (MAPE) WITH DIFFERENT WINDOW BASIS (*W*) AND VARIATION GRADE (*M*)

	Canada					Germany					Brazil					Luxembourg				
m	*w*					*w*					*w*					*w*				
	2	3	4	5	Average	2	3	4	5	Average	2	3	4	5	Average	2	3	4	5	Average
4	8.26	8.67	9.27	9.65	8.96	12.09	11.8	11.6	11.98	11.87	17.64	18.74	19.05	19.28	18.68	10.97	10.66	10.71	11.24	10.97
5	8.28	8.32	8.65	8.96	8.55	12.22	11.75	12.02	12.90	12.22	16.93	18.2	18.43	18.43	18.00	9.27	9.53	9.72	10.35	9.27
6	8.32	8.9	9.17	9.91	9.08	11.6	11.23	11.45	12.36	11.66	17.44	17.5	17.69	17.69	17.58	9.5	9.12	**9.11**	10.13	9.5
7	8.34	7.98	8.4	8.18	8.23	11.61	11.35	11.53	12.26	11.69	**13.58**	14.12	14.44	14.56	14.18	9.69	9.29	9.46	10.13	9.69
8	**7.02**	7.85	8.43	8.24	7.89	11.39	10.74	11.05	11.7	11.22	14.86	16.57	15.85	15.85	15.78	10.14	10.14	9.54	10.57	10.14
9	7.84	7.94	9.25	9.56	8.65	10.76	**10.71**	10.85	11.76	11.02	14.13	15.59	15.39	15.39	15.13	11.72	11.1	11.16	11.88	11.72
Average	8.01	8.28	8.86	9.08	8.56	11.61	11.26	11.42	12.16	11.61	15.76	16.79	16.81	16.87	16.56	10.22	9.97	9.95	10.72	10.22

	United Kingdom					Denmark					Nigeria					Yugoslavia				
m	*w*					*w*					*w*					*w*				
	2	3	4	5	Average	2	3	4	5	Average	2	3	4	5	Average	2	3	4	5	Average
4	7.76	6.89	7.15	7.15	7.24	**6.95**	**6.95**	**6.95**	**6.95**	6.95	21.29	21.7	22.77	23.73	22.37	40.96	44.27	46.56	46.56	44.59
5	7.11	7.28	7.52	7.52	7.36	7.46	8.03	8.33	8.33	8.04	15.46	15.46	15.46	16.5	15.72	37.84	37.61	40.45	40.45	39.09
6	7.2	7.47	7.71	7.9	7.57	7.98	8.28	8.52	8.52	8.33	18.73	19.85	19.03	19.28	19.22	37.63	39.68	44.84	44.84	41.75
7	6.97	7.38	7.68	7.68	7.43	8.68	8.28	8.68	8.68	8.58	16.82	18.33	17.64	17.85	17.66	33.62	36.93	42.07	40.94	38.39
8	8.27	8.16	**6.29**	6.34	7.27	7.17	7.21	7.78	7.78	7.49	14.06	14.06	14.54	17.34	15.00	37.02	35.28	36.32	36.32	36.24
9	8.49	8.23	7.38	7.4	7.88	7.92	8.07	8.22	8.05	8.07	14.43	14.43	**13.91**	15.46	14.56	**32.68**	40.68	44.47	44.47	40.58
Average	7.63	7.57	7.29	7.33	7.46	7.69	7.80	8.08	8.05	7.91	16.80	17.31	17.23	18.36	17.42	36.63	39.08	42.45	42.26	40.10

Note: Bold values denote the smallest forecast error (within *m*, *w*, by error measure).

TABLE 3

FUZZY TIME-SERIES FORECAST ACCURACY ($\hat{P}_1 + \hat{P}_2$) WITH DIFFERENT WINDOW BASIS (W) AND VARIATION GRADE (M)

	Canada					Germany					Brazil					Luxembourg				
m	w					w					w					w				
	2	3	4	5	Average	2	3	4	5	Average	2	3	4	5	Average	2	3	4	5	Average
4	1.08	1.54[a]	1.08	1.08	1.20	0.73	1.30[a]	1.30[a]	1.10	1.11	0.93	0.93	1.04	1.04	0.99	1.47[a]	1.47[a]	1.30[a]	1.30	1.39
5	0.46	0.79	0.46	0.33	0.51	0.70	1.07	0.90	0.50	0.79	1.29[a]	0.39	0.29	0.14	0.53	1.47[a]	1.47[a]	1.30[a]	1.10	1.34
6	1.17	0.83	1.17	1.29	1.12	1.13	1.13	1.50[a]	1.47[a]	1.31	1.21	0.93	0.79	0.64	0.89	1.47[a]	1.47[a]	1.47[a]	1.30[a]	1.43
7	0.83	1.42[a]	0.96	1.63[a]	1.21	0.57	0.73	0.73	0.90	0.73	1.36[a]	1.32	0.93	1.07	1.17	1.27	1.27	1.27	1.27	1.27
8	1.29	0.96	1.29	1.29	1.21	0.97	1.13	1.13	1.13	1.09	1.11	1.11	1.11	0.96	1.07	1.47[a]	1.47[a]	1.47[a]	1.47[a]	1.47
9	0.92	0.92	1.25[a]	0.92	1.00	1.13	0.57	0.57	1.17	0.86	1.36[a]	1.07	1.07	0.93	1.11	1.27	1.47[a]	1.47[a]	1.10	1.33
Average	0.96	1.08	1.04	1.09	1.04	0.87	0.99	1.02	1.05	0.98	1.21	0.96	0.87	0.80	0.96	1.40	1.44	1.38	1.26	1.37
	United Kingdom					Denmark					Nigeria					Yugoslavia				
m	w					w					w					w				
	2	3	4	5	Average	2	3	4	5	Average	2	3	4	5	Average	2	3	4	5	Average
4	0.63	1.63[a]	1.63[a]	1.75[a]	1.41	1.21	1.21	1.21	1.21	1.21	0.82	0.82	0.82	0.82	0.82	1.27	1.27	1.27	1.10	1.23
5	0.83	0.50	0.50	0.50	0.58	1.71[b]	1.71[b]	1.71[b]	1.71[b]	1.71	0.82	0.82	0.82	0.82	0.82	1.07	1.07	1.07	0.90	1.03
6	0.71	0.83	1.08	1.08	0.93	0.82	0.82	0.82	0.96	0.86	1.21	1.21	1.21	1.21	1.21	1.07	1.07	1.07	0.90	1.03
7	0.58	1.17	1.04	0.92	0.93	1.07	1.07	1.07	1.21	1.11	1.21	1.21	1.21	1.21	1.21	1.07	1.07	1.07	0.90	1.03
8	0.83	0.83	0.83	0.58	0.77	1.21	1.46[a]	1.46[a]	1.46[a]	1.40	0.82	0.82	0.82	0.82	0.82	1.07	1.07	1.07	0.90	1.03
9	1.17	0.71	1.08	0.96	0.98	0.96	0.96	0.96	1.21	1.02	1.21	1.21	1.21	1.21	1.21	1.23	1.07	1.07	0.90	1.07
Average	0.79	0.95	1.03	0.97	0.93	1.16	1.21	1.21	1.29	1.22	1.02	1.02	1.02	1.02	1.02	1.13	1.10	1.10	0.93	1.07

a. $\alpha \leq 0.05$.
b. $\alpha \leq 0.05$ and $\beta \geq 0.95$.

Downloaded from jtr.sagepub.com at Universita della Calabria on March 8, 2011

TABLE 4

GREY THEORY FORECAST ACCURACY

	Canada			Germany			Brazil			Luxembourg		
n	MAPE	RMSPE	$\hat{p}_1 + \hat{p}_2$	MAPE	RMSPE	$\hat{p}_1 + \hat{p}_2$	MAPE	RMSPE	$\hat{p}_1 + \hat{p}_2$	MAPE	RMSPE	$\hat{p}_1 + \hat{p}_2$
4	10.22	11.13	1.63[a]	12.57	15.72	0.57	15.9	20.16	1.04	8.85	12.01	1.47[a]
5	10.65	12.19	1.29	12.18	16.48	0.70	13.91	17.84	1.32[a]	11.78	15.30	1.47[a]
6	10.40	13.05	1.25[a]	12.36	16.97	0.93	16.44	21.94	1.36[a]	14.60	19.28	1.47[a]
7	11.07	14.6	0.71	12.88	17.97	1.10	19.50	26.59	1.36[a]	16.19	23.28	1.67[b]
	United Kingdom			Denmark			Nigeria			Yugoslavia		
n	MAPE	RMSPE	$\hat{p}_1 + \hat{p}_2$	MAPE	RMSPE	$\hat{p}_1 + \hat{p}_2$	MAPE	RMSPE	$\hat{p}_1 + \hat{p}_2$	MAPE	RMSPE	$\hat{p}_1 + \hat{p}_2$
4	9.31	11.48	1.04	9.96	12.66	0.82	12.42	15.68	1.32[a]	44.55	55.16	0.9
5	9.19	12.08	1.17	9.46	11.87	0.68	13.93	17.33	0.96	39.73	47.32	0.87
6	9.87	12.52	1.04	9.42	11.30	1.07	15.06	18.30	0.82	38.89	42.55	1.43[a]
7	11.05	13.90	0.38	10.95	13.09	1.36[a]	15.77	19.74	1.21	40.76	49.61	0.87

Note: MAPE = mean absolute percentage error; RMSPE = root mean squared percentage error.
a. $\alpha \leq 0.05$.
b. $\alpha \leq 0.05$ and $\beta \leq 0.95$.

TABLE 5

FORECASTING ACCURACY OF EACH MODEL ($N = 5$): ERROR AND RANK

Error measure	Models	Canada	Germany	Brazil	Luxembourg	United Kingdom	Denmark	Nigeria	Yugoslavia	Score
MAPE	Fuzzy	8.28(2)	11.26(1)	16.79(3)	9.97(1)	7.57(1)	7.80(1)	17.31(4)	39.08(2)	25
	Grey	10.65(4)	12.18(2)	13.91(2)	11.78(2)	9.19(2)	9.46(3)	13.93(1)	39.73(3)	21
	DMA	10.61(3)	12.81(3)	13.04(1)	13.54(3)	9.50(3)	9.54(4)	15.49(2)	42.29(4)	12
	DES	7.70(1)	13.4(4)	17.07(4)	15.64(4)	9.55(4)	8.79(2)	15.96(3)	38.08(1)	17
RMSPE	Fuzzy	9.22(1)	13.44(1)	21.34(3)	12.32(1)	10.10(1)	9.75(1)	19.37(4)	47.75(4)	24
	Grey	12.19(3)	16.48(2)	17.84(1)	15.30(2)	12.08(3)	11.87(4)	17.33(1)	47.32(2)	22
	DMA	13.32(4)	16.49(3)	18.26(2)	16.93(3)	12.31(4)	11.54(3)	18.40(3)	47.44(3)	15
	DES	9.22(1)	17.71(4)	21.75(4)	21.54(4)	10.71(2)	11.39(2)	17.74(2)	44.64(1)	20
$\hat{p}_1 + \hat{p}_2$	Fuzzy	1.08(2)	0.99(2)	0.96(4)	1.44(3)	0.95(2)	1.21(1)	1.02(1)	1.10(2)	23
	Grey	1.29(1)	0.70(4)	1.32[a](2)	1.47[a](2)	1.17(1)	0.68(4)	0.96(2)	0.87(4)	20
	DMA	0.83(3)	1.13(1)	1.21(3)	1.67[a](1)	0.50(3)	1.07(2)	0.82(3)	1.07(3)	21
	DES	0.83(3)	0.90(3)	1.46[a](1)	1.30[a](4)	0.50(3)	0.96(3)	0.57(4)	1.23(1)	18

Note: MAPE = mean absolute percentage error; RMSPE = root mean squared percentage error.
a. $\alpha \leq 0.05$.
b. $\alpha \leq 0.05$ and $\beta \leq 0.95$.

predicting the same years, the compared models were not necessarily fitted on the same length of series.

Tables 5 and 6 display each model's accuracy in terms of both the forecast error (MAPE, RMSPE, and $\hat{p}_1 + \hat{p}_2$) and the relative performance of each model within the country (rank), using two predetermined lengths of historical data series ($N = 5$ and $N = 7$). The first number in each cell is either MAPE or RMSPE. The rank (in parentheses) indicates how well the model forecasted that country's arrival figures in comparison to the other three models. For example, Canada's fuzzy figures (table 7, row 2, column 3) should be interpreted as follows: the MAPE of the fuzzy time-series model's prediction of Canada's figures, using 5 data points, was 8.28%. It ranked second, following the DES model's prediction in the first place with a MAPE of 7.70%. The score of each country is a number assigned to each rank as follows: rank 1 is assigned the score of 4, rank 2 is assigned the score of 3, rank 3 is assigned a score of 2, and rank 4 is assigned a score of 1. The total score of a method (the right column) is the sum of the scores of all eight countries. Thus, the larger the total score, the more accurate the model.

The results as presented in the two tables indicate that no method consistently outperformed the others. One model may outperform the others in forecasting arrivals from a single country but fail with other countries. For example, DES is the most accurate method (in terms of MAPE) for predicting arrivals from Canada but the least accurate in predicting arrivals from Germany. The fuzzy time-series model performs well for many countries but is the worst model when it comes to Nigeria. The order of performance (rank) within an individual country might be slightly affected by the choice of error magnitude measure, but the total score is not. Note also that within each country, the rank depends on the length of the series. When 5 years of previous data are used to fit the model, the fuzzy time series scores the highest (i.e., the most accurate forecast), the grey theory model is in second place, and DMA is the worst. When 7 years of data are used instead, the rank order (best to worst) changes to DES, fuzzy time series, grey theory

Downloaded from jtr.sagepub.com at Universita della Calabria on March 8, 2011

TABLE 6

FORECASTING ACCURACY OF EACH MODEL ($N = 7$): ERROR AND RANK

Error Measure	Models	Canada	Germany	Brazil	Luxembourg	United Kingdom	Denmark	Nigeria	Yugoslavia	Score
MAPE	Fuzzy	9.08(2)	12.16(2)	16.87(2)	10.72(1)	7.33(1)	8.05(2)	18.36(4)	42.26(3)	23
	Grey	11.07(4)	12.88(3)	19.5(4)	16.19(3)	11.05(4)	10.95(3)	15.77(2)	40.76(2)	13
	DMA	11.03(3)	13.61(4)	19.19(3)	16.96(4)	10.7(3)	11.16(4)	17.91(3)	44.77(4)	12
	DES	6.56(1)	11.25(1)	16.72(1)	13.12(2)	9.13(2)	7.87(1)	12.68(1)	33.27(1)	30
RMSPE	Fuzzy	10.08(2)	14.29(1)	21.06(1)	13.10(1)	9.88(2)	9.88(1)	21.72(4)	51.09(3)	23
	Grey	14.6(3)	17.97(3)	26.59(3)	23.28(3)	13.9(4)	13.09(3)	19.74(2)	49.61(2)	17
	DMA	15.29(4)	18.33(4)	26.88(4)	24.47(4)	13.28(3)	13.62(4)	20.63(3)	52.65(4)	10
	DES	7.74(1)	14.37(2)	21.54(2)	15.85(2)	9.83(1)	10.13(2)	14.86(1)	42.38(1)	28
$\hat{p}_1 + \hat{p}_2$	Fuzzy	1.09(1)	1.05(2)	0.80(3)	1.26(4)	0.97(1)	1.29(3)	1.02(4)	0.93(2)	20
	Grey	0.71(4)	1.10(1)	1.36[a](2)	1.67[b](1)	0.38(3)	1.36[a](2)	1.21(1)	0.87(4)	22
	DMA	0.96(2)	0.77(3)	0.79(4)	1.50[a](2)	0.25(4)	1.61[a](1)	1.11(2)	1.27(1)	21
	DES	0.92(3)	0.67(4)	1.57[a](1)	1.47[a](3)	0.96(2)	1.36[a](2)	1.07(3)	0.90(3)	19

Note: MAPE = mean absolute percentage error; RMSPE = root mean squared percentage error.
a. $\alpha \leq 0.05$.
b. $\alpha \leq 0.05$ and $\beta \geq 0.95$.

TABLE 7

THE BEST FORECASTING ACCURACY OF EACH MODEL: ERROR AND RANK

Error Measure	Models	Canada	Germany	Brazil	Luxembourg	United Kingdom	Denmark	Nigeria	Yugoslavia	Score
MAPE	Fuzzy	7.02(2)	10.71(1)	13.58(2)	9.11(2)	6.29(1)	6.95(1)	13.91(4)	32.68(2)	25
	Grey	10.22(3)	12.18(3)	13.91(3)	8.85(1)	9.19(3)	9.42(3)	12.42(2)	38.89(3)	19
	DMA	10.61(4)	12.81(4)	13.04(1)	9.45(3)	9.19(3)	9.54(4)	12.23(1)	42.29(4)	16
	DES	6.02(1)	11.25(2)	16.11(4)	12.8(4)	8.93(2)	7.22(2)	12.62(3)	32.38(1)	21
RMSPE	Fuzzy	8.22(2)	12.75(1)	17.36(1)	11.16(1)	7.99(1)	9.1(1)	15.2(2)	41.07(1)	30
	Grey	11.13(3)	15.72(3)	17.84(2)	15.3(3)	12.08(4)	11.3(3)	15.68(3)	42.55(3)	16
	DMA	11.58(4)	16.12(4)	18.26(3)	11.79(2)	11.44(3)	11.54(4)	15.71(4)	47.44(4)	12
	DES	7.51(1)	14.37(2)	21.51(4)	15.78(4)	9.76(2)	9.45(2)	14.86(1)	42.28(2)	22
$\hat{p}_1 + \hat{p}_2$	Fuzzy	1.54[a](2)	1.50[a](1)	1.36[a](2)	1.47[a](2)	1.75[a](1)	1.71[b](1)	1.21(2)	1.23(3)	26
	Grey	1.63[a](1)	1.10(3)	1.36[a](2)	1.67[b](1)	1.17(2)	1.36[a](3)	1.32[a](1)	1.43[a](1)	26
	DMA	0.96(4)	1.13(2)	1.71[b](1)	1.67[b](1)	0.96(3)	1.61[b](2)	1.11(3)	1.27(2)	22
	DES	1.42[a](3)	1.10(3)	1.71[b](1)	1.67[a](1)	1.75[a](1)	1.36[a](3)	1.07(4)	1.23(3)	21

Note: MAPE = mean absolute percentage error; RMSPE = root mean squared percentage error.
a. $\alpha \leq 0.05$.
b. $\alpha \leq 0.05$ and $\beta \leq 0.95$.

model, and DMA. The difference among the rank score of the four models is very small when forecasts accuracy are measured in terms of directional change, and the order is different when different lengths of previous data are used.

Table 7 compares the accuracy of the four models based on the second approach, that is, each model's best performance across the different time-series lengths is presented. Very much like the results of the first approach, none of the four methods consistently outperforms the others, and the choice of error measure has no considerable effect on the order (total score). Overall, the order of performance changes to fuzzy time series, DES, grey theory model, and DMA.

For the directional change, the value of $\hat{p}_1 + \hat{p}_2$, the difference of overall rank score among the four models is still small, although there is a difference among the four models for some individual countries.

Fuzzy time series seems to perform better with shorter data sets, and DES seems to be more accurate when more data are available (note, however, that the differences in the series lengths are very small—5 compared to 7—and both are considered rather short for forecasting purposes). In general, then, the forecast accuracy of the fuzzy time-series model and DES seems better than that of the grey theory model and DMA when predicting annual arrivals to the United States. This conclusion, however, does not hold true when the differences are tested for statistical significance using Wilcoxon's signed rank test.

Wilcoxon's Signed Rank Test

Table 8 outlines the results of 64 Wilcoxon tests with two predetermined lengths of data series (2 AI Models × 2 Benchmarks × 2 Series Lengths × 8 Countries of Origin). The critical T for a confidence level of 95% (a sample size of 11) is 10. That is, if $T(+)$ is smaller than 10, the difference between the benchmark model and the tested model is considered statistically significant. As clearly demonstrated in table 8, all of the $T(+)$ values for models fitted on five observations, and all but two for models fitted on seven observations, are larger than the critical T. That is, 62 out of 64 statistical tests indicate that the observed differences in forecast accuracy are not statistically significant. Moreover, of

Downloaded from jtr.sagepub.com at Universita della Calabria on March 8, 2011

TABLE 8

RESULTS OF THE WILCOXON TEST (T[+])

N	Benchmark Model	Tested Model	Canada	Germany	Brazil	Luxembourg	United Kingdom	Denmark	Nigeria	Yugoslavia
5	DMA	Fuzzy	22	24	29	16	28	13	30	18
		Grey	31	28	33	15	18	31	24	18
	DES	Fuzzy	32	33	22	15	16	22	26	23
		Grey	15	24	26	23	31	30	27	25
7	DMA	Fuzzy	26	34	21	23	14	11	29	24
		Grey	29	23	30	27	30	29	21	14
	DES	Fuzzy	25	22	25	22	10*	19	24	31
		Grey	17	28	27	23	27	9*	26	22

*Statistically significant ($p < .05$).

TABLE 9

RESULTS OF THE WILCOXON TEST (T[+])

Country	Fuzzy	Grey	DMA	DES	Best Model (MAPE)
Canada	24	9*	8*		DES
Germany		26	24	26	Fuzzy
Brazil	30	33		26	DME
Luxembourg	31		28	20	Grey
United Kingdom		18	12	12	Fuzzy
Denmark		14	13	31	Fuzzy
Nigeria	24	31		31	DME
Yugoslavia	28	19	15		DES

*Statistically significant ($p < .05$).

TABLE 10

RESULTS OF WILCOXON'S SIGN-RANK TEST (T[+])

Country	Fuzzy	Grey	DME	DES	Best Model (RMSPE)
Canada	24	9*	8*		DES
Germany		27	29	30	Fuzzy
Brazil		26	30	26	Fuzzy
Luxembourg		31	31	16	Fuzzy
United Kingdom		8*	12	12	Fuzzy
Denmark		14	13	31	Fuzzy
Nigeria	28	31	31		DES
Yugoslavia		22	14	30	Fuzzy

*Statistically significant ($p < .05$).

the two statistically significant tests, one indicates that the fuzzy time-series model outperforms a benchmark model (DES) in predicting the number of tourist arrivals (from the United Kingdom), but the second (Denmark) shows that the forecast of a benchmark model (DES) is more accurate than that of the grey theory model.

Tables 9 and 10 (MAPE and RMSPE, respectively) outline the results of the Wilcoxon test, comparing the best performance of the four models (across the different series lengths) to the predictions of the three benchmarks. Three paired predictions are tested within each country of origin. Out of 48 comparisons (2 Error Measures × 3 Paired Models × 8 Countries), only 5 are statistically significant. The DES prediction of arrivals from Canada is significantly more accurate than that of the DMA and grey theory predictions (both MAPE and RMSPE). In addition, the fuzzy time-series prediction of arrivals from the United Kingdom was significantly more accurate than that of the grey theory (RMSPE).

In summary, the 112 statistical tests clearly indicate that none of the forecasting methods outperforms the others consistently—the vast majority of the measured differences were not deemed statistically significant. Although in several cases fuzzy time series and DES seemed to have better forecasting accuracy, these differences were rarely statistically significant. Hence, this study does not support the notion that a forecast of annual tourist arrivals generated by the two AI models (fuzzy time series and grey theory), even when few data are available for model fitting, is likely to be more accurate than that of the traditional, simple-to-apply DMA and DES models.

SUMMARY AND CONCLUSION

Fuzzy time series and grey theory are two AI forecast methods that apply linguistic variables and accumulated generating operations (AGO) instead of the original data. Designed to handle short time series, the two seem most adequate for forecasting annual tourist arrivals, as the tourism literature has recently suggested.

The purpose of this study was to further explore the usefulness of the two methods in predicting annual tourist arrivals. Both AI methods are rather complicated and are not yet available as part of statistical/forecasting software. As such, they require rather skilled forecasters. Given the complexity and cost associated with applying these two AI methods, it is imperative that their performance be carefully compared with the accuracy of more traditional, and much easier, methods of forecasting.

DES and DMA were selected as the benchmarks. Both use the original time series, are commonly used in tourism-forecasting applications because of their simplicity and solid performances, and seemed most adequate because of the trend patterns in the data. The accuracy of the two AI models and the two benchmark models was tested using the number of annual tourist arrivals to the United States from eight

Downloaded from jtr.sagepub.com at Universita della Calabria on March 8, 2011

countries throughout a period of 18 years (1985-2002). Using rolling forecasts and holdout samples, the four models were fitted on series of various lengths, and the accuracy of their predictions was assessed using two error measures: MAPE and RMSPE. Wilcoxon's matched-pairs, signed-ranked, nonparametric test was used to evaluate the statistical significance of the observed differences in accuracy.

None of the models outperformed the others consistently, and most of the differences were not statically significant. Similarly, the forecasts of the two AI models were not significantly different from those of the benchmarks. This result is consistent with the conclusion drawn by the *M*-series competitions (Makridakis and Hibon 2000).

The findings of this study indicate that the complicated models are not likely to generate a more accurate forecast of annual tourist arrivals compared with the simple, traditional models. Given the notable cost associated with the application of these two AI forecasting methods, our finding of no significant accuracy advantage suggests that tourism forecasters should not rush to adopt these two AI methods without careful consideration. That these two forecasting methods were designed to perform better when only a few past observations are available does not necessarily indicate their ability to outperform very simple methods of forecasting annual tourism demand.

An obvious limitation of this study is that, although larger than previous studies, the eight annual time-series sample chosen from a framework of 87 countries of origin may still be too small. Furthermore, this study's attempt to cover a wide range of data patterns resulted in the inclusion of tourist arrival data from countries such as Nigeria and Yugoslavia and the exclusion of larger countries of origin such as France, Italy, Ireland, Australia, and New Zealand. The performance of the tested forecasting models with these excluded countries of origin might be of more interest to practitioners because these countries generate more annual tourist traffic. Future research may extend this study by testing additional data. In addition, the data used in this study were all arrivals to the same destination. Future efforts, therefore, might include data of arrivals to multiple destinations. A second venue for future research concerns the data pattern. Although the two AI models did not outperform the two simple models, the results might change if the data are categorized according to patterns. Some of the tested time series contained abrupt changes, violating the models' assumption. Accuracy tests that control for the pattern, particularly abrupt changes, might prove that some models outperform others in certain pattern-related situations. Finally, other extensions of this work might attempt to improve the accuracy of these models (both the AI and the benchmarks). Possible venues include using the Markov chain (Hsu and Wen 1998), applying adaptive fuzzy methods (Lin, Su, and Hsu 2001) to improve the basic grey theory model, and combining the forecasts of different models.

REFERENCES

Au, N., and R. Law (2000). "The Application of Rough Sets to Sightseeing Expenditures." *Journal of Travel Research*, 31 (1): 70-77.

Box, G. (1991). "Understanding Exponential Smoothing: A Simple Way to Forecast Sale and Inventory." *Quality Engineering*, 4 (1): 143-51.

Chiang, J. S., P. L. Wu, S. D. Chiang, T. J. Chang, S. T. Chang, and K. L. Wen (1998). *Introduction of Grey Theory*. Taiwan: GAO-Li.

Daniel, W. W. (1990). *Applied Nonparametric Statistics*. 2nd ed. Duxbury Advanced Series in Statistics and Decision Sciences. Boston: PWS-KENT.

Deng, J. L. (1982). "Control Problem of Grey System." *Systems and Control Letters*, 1:288-94.

Frechtling, D. C. (2001). *Forecasting Tourism Demand: Methods and Strategies*. Oxford: Butterworth-Heinemann.

Goh, C., and R. Law (2003). "Incorporating the Rough Sets Theory into Travel Demand Analysis." *Tourism Management*, 24 (5): 511-18.

Henriksson, R. D., and R. C. Merton (1981). "On Market Timing and Investment Performance 2: Statistical Procedures for Evaluating Forecasting Skills." *Journal of Business*, 54:513-33.

Hsu, L. C. (2003). "Applying the Grey Prediction Model for the Global Integrated Circuit Industry." *Technological Forecasting and Social Change*, 70 (6): 563-74.

Hsu, C. I., and Y. U. Wen (1998). "Improved Grey Prediction Models for Trans-Pacific Air Passenger Market." *Transportation Planning and Technology*, 22:87-107.

Hwang, J. R., S. M. Chen, and C. H. Lee (1998). "Handling Forecasting Problems Using Fuzzy Time Series." *Fuzzy Sets and Systems*, 100:217-28.

Law, R. (2000). "Back-Propagation Learning in Improving the Accuracy of Neural Network-Based Tourism Demand Forecasting." *Tourism Management*, 21:331-40.

Law, R., and N. Au (1999). "A Neural Network Model to Forecast Japanese Demand for Travel to Hong Kong." *Tourism Management*, 20:89-97.

Lim, C., and M. McAleer (2001). "Forecasting Tourist Arrivals." *Annals of Tourism Research*, 28 (4): 965-77.

Lin, C. B., S. F. Su, and Y. T. Hsu (2001). "High-Precision Forecast Using Grey Models." *International Journal of Systems Science*, 32 (5): 609-19.

Makridakis, S., and M. Hibon (2000). "The M3-Competition: Results, Conclusions and Implications." *International Journal of Forecasting*, 16:451-76.

Martin, C. A., and S. F. Witt (1989). "Forecasting Tourism Demand: A Comparison of the Accuracy of Several Quantitative Methods." *International Journal of Forecasting*, 5:7-19.

Mincer, J., and V. Zarnowitz (1969). "The Evaluation of Economic Forecasts." In *Economic Forecasts and Expectations*, edited by J. Mincer. New York: National Bureau of Economic Research.

Office of Travel and Tourism Industries (1996). "Historical Arrivals to the United States (1985-1995), Summary of the International Travel to the United States." http://tinet.ita.doc.gov (accessed March 23, 2004).

Office of Travel and Tourism Industries (2004). "International Arrivals to the U.S.: Historical Visitation 1996-2002." U.S. Department of Commerce. http://tinet.ita.doc.gov (accessed March 23, 2004).

Schwartz, Z., and E. Cohen (2004a). "Subjective Estimates of Occupancy Forecasts Uncertainty by Hotel Revenue Managers." *Journal of Travel and Tourism Marketing*, 16 (4): 59-66.

——— (2004b). "Hotel Revenue Management Forecasting: Evidence of Expert Judgment Bias." *Cornell Hotel and Restaurant Administration Quarterly*, 45 (1): 85-98.

Schwartz, Z., and S. Hiemstra (1997). "Improving the Accuracy of Hotel Reservations Forecasting: Curves Similarity Approach." *Journal of Travel Research*, 36 (1): 3-14.

Snyder, R. D., A. B. Koehler, and J. K. Ord (2002). "Forecasting for Inventory Control with Exponential Smoothing." *International Journal of Forecasting*, 18 (1): 5-18.

Song, H., and S. F. Witt (2000). *Tourism Demand Modeling and Forecasting: Modern Econometric Approaches*. Oxford: Pergamon.

——— (2003). "Tourism Forecasting: The General-to-Specific Approach." *Journal of Travel Research*, 42 (1): 65-74.

Song, H., S. F. Witt, and T. C. Jensen (2003). "Tourism Forecasting: Accuracy of Alternative Economic Models." *International Journal of Forecasting*, 19:123-41.

Song, Q., and B. S. Chissom (1993). "Forecasting Enrollments with Fuzzy Time Series: Part I." *Fuzzy Sets and Systems*, 54:1-9.

——— (1994). "Forecasting Enrollments with Fuzzy Time Series: Part II." *Fuzzy Sets and Systems*, 62:1-8.

Tashman, L. J. (2000). "Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review." *International Journal of Forecasting*, 16:437-50.

Turner, L., and S. F. Witt (2001). "Forecasting Tourism Using Univariate and Multivariate Structural Time Series Models." *Tourism Economics*, 72:135-47.

Wang, C. H. (2004). "Predicting Tourism Demand Using Fuzzy Time Series and Hybrid Grey Theory." *Tourism Management*, 25 (3): 367-74.

Weatherford, L. R., and S. E. Kimes (2003). "A comparison of Forecasting Methods for Hotel Revenue Management." *International Journal of Forecasting*, 19:401-15.

Wheelwright, S. C., and S. Makridakis (1980). *Forecasting Methods for Management*. New York: John Wiley.

Witt, S. F., and C. A. Witt (1992). *Modeling and Forecasting Demand in Tourism*. London: Academic Press.

Witt, S. F., H. Song, and P. Louvieris (2003). "Statistical Testing in Forecasting Model Selection." *Journal of Travel Research*, 42 (2): 151-58.

Wong, K. F. (2002). "Introduction: Tourism Forecasting State of the Art." *Journal of Travel and Tourism Marketing*, 13 (1/2): 1-3.

Downloaded from jtr.sagepub.com at Universita della Calabria on March 8, 2011