Special Issue: The Economics of Revenue Management in Hospitality and Tourism

TOURISM ECONOMICS
The Business and Finance of Tourism and Recreation

Tourism Economics
1–19

DOI: 10.1177/1354816619884800
journals.sagepub.com/home/teu
SAGE

Enhancing the accuracy of revenue management system forecasts: The impact of machine and human learning on the effectiveness of hotel occupancy forecast combinations across multiple forecasting horizons

Zvi Schwartz
University of Delaware, USA

Timothy Webb
University of Delaware, USA

Jean-Pierre I van der Rest
Leiden University, the Netherlands

Larissa Koupriouchina
Hotelschool The Hague, the Netherlands

Abstract
Reporting on three separate studies in the context of hotel revenue management systems, this article explores the interaction between two established methods of accuracy enhancement—forecast combinations and learning. In line with theoretical considerations, our empirical investigation suggests that as learning occurs, the capacity of combinations to improve forecast accuracy diminishes in scenarios where the combined elements are independent of each other. Conversely, in the more realistic typical scenario of user overrides of system forecasts, where the elements of the combinations are dependent, the learning-driven efficacy of forecast combinations appears to vary across forecasting horizons. We find no impact of learning on combination effectiveness in the

Corresponding author:
Zvi Schwartz, Hospitality Business Management, Alfred Lerner College of Business & Economics, University of Delaware, 14W. Main Street, Raub Hall, Newark, DE 19716, USA.
Email: zvi@udel.edu

shorter forecasting horizons of 21 days or less and a surprisingly positive impact in the longer horizons. This counterintuitive finding has important practical implications for hotel revenue management practices.

Keywords
Combination, forecasting, hotel, learning, revenue management

Introduction

The accuracy of daily demand forecasting is a crucial performance element of hotel revenue management systems (RMSs) (Cross et al., 2009; Haensel and Koole, 2011; Talluri and van Ryzin, 2006; Weatherford and Kimes, 2003). In recent years, these systems have been challenged by several factors, including the lodging industry's ongoing transition into the multifaceted environment of total hotel revenue management (RM) practices (e.g. Fairly, 2013), the flexing booking window (Dacko, 2004; Forgacs, 2010; Mayock, 2009, 2011, Webb, 2016), and the increase in data analytics informed practices. The word on the street is that "revenue managers are getting more and more responsibility in forecasting future performance, both in revenue and profit" (HotelsNewsNow, July 1, 2019).

The move toward total RM increases the potential damage caused by forecast inaccuracy. For example, for a non-room revenue-generating center in the hotel, where the volume depends (in part or completely) on rooms' occupancy, the optimization algorithms are based on the same room occupancy predictions used by the rooms division. So, the same forecast error is now causing a larger potential damage through suboptimal RM decisions (rooms allocation) and with the other area(s) in the hotel that rely on these forecasts. As for the booking window, it appears to flex with economic cycle and advances in information technology. However, not all forecasting methods work equally across the various sizes of booking windows, or with windows that shrink, and expand, faster than they used too. Research is yet to identify the best forecasting methods to address this challenge. Consequently, hotel RM research continues to explore ideas on how to improve forecasting accuracy and this study continues this line of research.

This article focuses on the interaction of two, well-established, forecasting accuracy enhancement processes: combinations and learning. For several decades now, the literature on forecasting has demonstrated that one could generate better accuracy by combining forecasts (e.g. Timmerman, 2006). Forecast combination refers to the practice of generating a single forecast when multiple forecasts for the same variable exist. As discussed in more details in the next section, there are multiple theoretical reasons why it is a good practice to combine forecasts, and there are several methods in which forecasters could combine these multiple predictions of the same variable. Considerable progress was also gained in the area of machine and human learning (expertise levels), and how learning enhances forecasting accuracy through training and systematic feedback (e.g. Collopy et al., 2001). In this context, learning refers to the forecast performance, and more specifically to improvements in forecast accuracy (i.e., lower forecast errors) driven by experience gained over time.

This article's main contribution is being first to suggest, and to empirically explore, whether there is an interplay between these two well-established accuracy-enhancing methods of combinations and learning. The article starts with a review of the literature on forecast combinations and learning and provides a theoretical explanation for why an interaction effect is hypothesized. We then assess empirically the impact of one method (learning) on the second (combination) in the context of hotel RM. Our empirical investigation is comprised of three separate studies: two preliminary explorations and a main, large-scale study. The three studies contrast data from three distinctly different settings. The first study uses a 60-week "novice" data set from a typical academic classroom environment, where a systematic learning and practicing of quantitative forecasting methods occurred. The second preliminary study explores a professional environment, using a single hotel's RM team forecasts over a period of 2 years, where, uncharacteristically, machine and human forecasts were generated independently of each other. Lastly, the main study analyzes a very large data set, with 900 hotels, 694 forecasters, generating 1322 unique hotel–employee pairs. The data set consists of millions of daily occupancy forecasts, where forecast combinations take the form of user (expert) overrides. User override of machine forecasts is a very common way in which algorithmic (machine) and subjective forecasts are combined in hotels that use RM software (e.g. Arora, 2018; Revenue Analytics, 2018). All three studies assess whether learning and accuracy improvement occurred over time and explore how learning affected the accuracy improvement potential of the combination methods. The article concludes by exploring the theoretical and practical implications of the empirical findings, discussing the study's limitation, and suggesting venues for future research.

Background: Forecast combinations, expert overrides, and machine and human learning

Research has made a substantial contribution to the theory and practice of forecasting since the early 1960s (e.g. Fildes et al., 2008). A notable outcome of this extensive research effort is the recommendation to combine forecasts. It was concluded that forecasting performance can be substantially improved by combining multiple individual forecasts (Bates and Granger, 1969; Clemen, 1989). Armstrong (1986: 96) maintains that combination of forecasts use more information and produce more accurate forecasts by reducing errors, cheating, and unreasonable assumptions.

This advantage of forecasting combination is well supported by empirical evidence. Reporting on the M3-competition, Makridakis and Hibon (2000: 458) confirm that forecast combinations of various methods outperform, on average, the methods being combined and other methods. Armstrong (2001a) examined 30 empirical studies published between 1938 and 2001. He found that combining forecasts ex ante reduced forecasting error on average by 12.5% (ranging from 3% to 24%) and that the more the data and methods differ, the greater the expected improvement in forecasting accuracy. Utilizing data from the M3-competition, Hibon and Evgeniou (2005: 23) nuance these conclusions. They find that (1) if one always uses the same method or same combination for all time-series, then the best individual methods and combinations perform similarly. However, the worst performance among the individual methods is significantly worse than the worst performance among the combinations; (2) the performance of combinations drops significantly as more methods are combined; and (3) choosing an individual method out of a set of available methods is riskier than choosing a combination, and so, in practice, overall the chosen individual method may have significantly worse performance than the chosen combination.

The forecasting literature discusses a wide range of methods to combine independently generated forecasts (Clemen, 1989; Lahiri et al., 2013; Newbold and Harvey, 2002; Timmermann, 2006): From a simple average, where the weight for each individual forecast is determined by expert judgment (e.g. Collopy and Armstrong, 1992), through Bayesian model averaging (e.g.

Hoeting et al., 1999), to least squares regression (e.g. Granger and Ramanathan, 1984). Interestingly, empirical studies generally find that simple combination schemes (e.g. equal weight) most often outperform more complex weighted combination strategies (Smith and Wallis, 2009).

Several notable hospitality and tourism studies provide domain-specific supporting evidence to the accuracy improvement potential of combinations, as well as insight on combination schemes (e.g. Chu, 1998; Shen et al., 2008, 2010; Song et al., 2009; Wong et al., 2007). Expert's subjective adjustments of model predictions have been observed as a common way in which occupancy forecasts are combined in hotels where RMSs are used. User overrides of machine predictions is a "must" feature in all of the leading hotel RM computerized systems, but research is yet to address the many aspects related to these user overrides. The general literature on such expert adjustments suggest two theoretical scenarios in which overrides improve the quality of the forecast. In chapter 2 of a book on the topic, Franses (2014) provides in details the mathematical notations and summarizes (p. 17) the two scenarios as follows:

> The first angle originates from the possibility that a forecast from an econometric model or statistical algorithm may lead to a forecast error of which some part can be expected by an expert: for example, due to a somehow foreseeable regime shift or a sudden change in one of the explanatory variables. The second angle is associated with the prediction errors from the expert-adjusted forecast, which can be shown to be minimized (using a squared loss function) when an expert adds a truly relevant term to the model forecast.

Another relevant forecast accuracy research insight is related to learning, training, and feedback. Referring to Collopy et al. (2001), and to Stewart (2001) work on human forecasters learning, Fischhoff (2001) argues that forecasters gain expertise and become skilled in their particular domain by training, having direct experience and researching their topic.

Feedback, in various forms, is an important element in the process of forecasting learning and improvement, as it impacts forecast accuracy in single (e.g. Boweden, 1989) and recurrent forecasting tasks (Rieg, 2008). For example, Remus et al. (1996) found that task information feedback generated significantly better forecasting performance than performance outcome feedback and that adding cognitive information feedback to task information feedback did not significantly improve forecasting performance. Learning and feedback are important in the hotel RM forecasting setting as well. While human judgment is an indispensable element of even the most computerized systems, research indicates that these forecasting judgments within the hospitality RM domain are prone to bias and, as such, could benefit from feedback and systematic learning/improvement processes. For example, forecast overriding decisions by hotel managers are influenced by perceptions of how they are assessed by higher levels of management (Bendoly, 2013), judgmental forecasting accuracy was found to be unrelated to perceived forecasting uncertainty by revenue managers (Schwartz, 2003), and the nature of the user interface was found to influence how revenue managers override machine-based forecasts (Schwartz and Cohen, 2004). Moreover, Koupriouchina et al. (2014) reported that hotel revenue managers might not fully understand the complex nature and inherent difficulties of forecasting error measurements. Their findings were supported by Pereira (2016) who found inconsistencies in the results of forecasting measures and found that different accuracy measures produced contradictory answers, not only within a specific hotel's segment but also across segments. Accordingly, Mukhopadhyay et al. (2007) suggested a systematic feedback process for airline revenue management analysts, designed to increase the accuracy of their unconstrained demand forecast-adjustments and reduce their adjustment bias.

Learning and feedback occur with machine-based forecasting as well, where learning refers to computerized systems that can learn from data. More specifically, machine learning is about modeling the relation between a set of input variables and one or more output variables using a finite set of observations and it requires a theoretical framework to automatically estimate a suitable model and parameters. Machine learning in forecasting often uses neural networks, and it has attracted considerable attention in the hospitality and tourism literature (e.g. Law, 1998, 2000; Padhi and Aggarwal, 2011; Zheng et al., 2012). Interestingly, it is even applied to determine forecast combination weights (Prudêncio and Ludermir, 2005; Yang, 2004) using either adaptation (i.e. trying to be as good as the best forecast) or improvement (i.e. trying to significantly outperform each individual forecast).

To summarize, our review of the literature established that two important methods of accuracy improvement are forecast combinations and learning. As explained, our study aims to explore how these two methods of accuracy improvement interact. More specifically, it seeks to understand how progress (accuracy improvement) with one method (learning) impacts the ability of the second one (forecast combination in its various forms) to improve the forecast accuracy.

Both theoretical and empirical considerations suggest that learning and training of forecasters and machines might have an impact on the effectiveness of forecast combination. Timmerman's (2006) review of the forecasting combination literature listed four possible theoretical explanations as to why forecasts combinations improve accuracy:

1. Individuals (different forecasters or forecasting methods) have "private" or unobserved information regarding the independent variables (IVs) (Bates and Granger, 1969; Clemen, 1987, as cited in Timmerman 2006: 138).
2. Individuals are differently effected by, or respond with a different speed to, structural breaks over time (e.g. Timmerman, 2006).

Moreover, forecast combinations reduce the damaging impact of

3. misspecification bias of unknown form (Clemen, 1989; Diebold and Lopez, 1996; Makridakis,1989; Stock and Watson, 2001, 2004, as cited in Timmerman, 2006: 138)
4. bias introduced by individual preferences for a different loss function (Christoffersen and Diebold, 1997; Zellner, 1986, as cited in Timmerman, 2006: 139)

The first and the third theoretical explanations above appear to suggest that learning and feedback might reduce the motivation for forecast combinations. Learning over time could mean that individual information becomes "less private" as forecasters learn about possible important variables that other forecaster, or other methods, have used successfully. This learning could happen through trial and error or through information sharing processes. Regardless of the learning method, if the outcome is indeed a reduction in relevant information becoming less private and more shared, forecast combinations' potential to improve accuracy might diminish. Similarly, learning, and specifically feedback, might reduce bias of unknown form. As humans and machines repeatedly assess their forecast accuracy and make efforts to improve, the outcome might be a better understating of underlying patterns in the forecasting process, including acknowledgment of existing biases, and taking adequate corrective actions. Again, it follows that if the learning, training, and feedback lead to a reduction in the harmful impact of unknown bias, then it will diminish the need to use forecasting combinations. Moreover, empirical evidence suggests that forecast combinations work best when there is high uncertainty about the situation (past and future), about which forecasting method is best, and when there are large errors to be avoided (Armstrong, 2001b). Obviously, as learning, training, and feedback occur, the conditions listed above are bound to lessen, as the forecasters (or the machines) develop familiarity with the situation, reduce uncertainty, and reduce the size of the forecasting errors.

To summarize, both theoretical and empirical considerations appear to suggest that learning could diminish the role of combinations. That is, the effectiveness, and consequently the desirability, of the more costly and cumbersome forecast combination method diminishes when learning occurs. Accordingly, this study hypothesizes the following:

> **H1:** When learning occurs, the capacity of forecast combinations to improve the accuracy diminishes.

The next part of the article empirically explores this theoretical conjecture in the context of hotel RM forecasting. As alluded to earlier, we explore three cases. The first two are exploratory in nature. The third, full-scale study, is the focus of this article. Accordingly, the study used three separate data sets, each representing distinct hotel occupancy forecasting circumstances and different learning and forecast combination characteristics.

Study 1

Methodology

The first data set is associated with a novice-forecasting context. It was adopted because it is better suited to assess the phenomenon in an environment where learning and training in the art of forecasting was a purposeful, premeditated, controlled, and closely monitored educational process in a university course. While this is not a typical circumstance in industry practices, it does occur sometimes with employees in a training period. We used a student sample from a Mid-Atlantic large university in the United States, which utilized real-time hotel occupancy forecasting as a key classroom activity in a specialized hospitality RM course. Over five semesters (five classes over 2.5 years), different student teams forecasted the average weekly occupancy of the local hotel market, one week ahead, generating 14 weeks of observations in each of the five semesters. As part of the student's learning experience, each week, the forecasts were compared to the actual occupancy in the market, that is, to figures provided weekly by a third party (STR Global Inc.). Over the first 5 weeks of each semester, the students were instructed on formal, quantitative forecasting techniques such as *Moving Average*, *Exponential Smoothing*, and *Holt–Winters Exponential Smoothing*, among others. Each week the students were presented with the actual occupancy figures as soon as they were published by STR. Feedback to the students was provided in the format of forecast error (Mean Absolute Error and Mean Absolute Percentage Error) and relative ranking of their team's accuracy performance compared to the other competing forecast teams. A classroom discussion on what methods had been used by the various teams followed, exploring reasons for inaccuracies, as well as strategies for improvements in the following week's round of forecasts. To summarize, the forecasting learning and training for this group of participants included not only exposure to, and practice of, new quantitative forecasting models but also the systematic application of accuracy feedback, identifying and reducing bias, incorporating exogenous/private information, and making subjective adjustments. The first forecast and the last forecast of each semester were excluded from the data since the students were often confused about expectations in the first week of the forecast competition and tended to lose interest during the last week of the semester. Taking this into consideration, each group within each semester provided 11 weeks of observations. In the preliminary phase of the data analysis, each group forecast was evaluated against the actual occupancy percentage to obtain an error assessment of accuracy.

Forecast combinations. Each group's weekly forecast was averaged with each other group's forecast to create a two-team combination forecast. For instance, if there were three groups in a certain week round, there would be a total of six forecasts for that week: three forecasts individually and three pairwise combinations. All possible pairwise forecast combinations were generated using a computer program (VBA© 2016) written for this project. We used equal weights combinations for two reasons. First, studies suggest that equal weight combinations are generally a safe choice because they rarely deliver poor performance and that overall they seem to be as effective as unequal weights (Aiolfi et al., 2011; Genre et al., 2013; Goodwin, 2009; Hibon and Evgeniou, 2005). Furthermore, equal weights are considered optimal when the individual forecast errors have identical variance and identical pair-wise correlations—a situation likely to occur when the models are based on similar data and perform roughly the same (Timmermann, 2006). This appears to be the case with the students learning data set. As with the individual forecasts, the combined predictions were contrasted with the actual occupancy of the hotel market for that week. In total (individual and combined predictions), the student data set consisted of 1352 observations, that is, 1352 initial forecast errors.

Regression model. The dependent variable (DV) used was the forecast accuracy. Specifically, we tested the *Absolute Percentage Error* (APE) measure, calculated as the absolute difference between the weekly occupancy forecast and the actual occupancy for that week, divided by the actual weekly occupancy. The independent variables included the week the forecast was made (ranging from 2 to 12); the forecast combination status (a categorical variable with zero denoting a single group forecast and one denoting a two-team combined forecast); and an interaction term between the given week (the time variable) and the combination indicator. The time (week) variable was designed to assist in assessing whether learning occurs and capture the learning and improvement of the student teams over time. This was done by categorizing each forecast into the period that the forecast was made—in the case of the first data set, it was the week period. The first week was coded as 2, sequentially increasing the week variable by one for each proceeding week. The interaction term was calculated by multiplying the variables week and the combination status, that is, Week $\times$ Combination. The final data set consisted of 137 unique individuals or pairwise comparisons measured over a possible 11 weeks. The cross-sectional time series data required panel model estimation. Several tests were used for model specification, most notable the Hausman test was found insignificant in all instances inferring that the random effects model was most appropriate.

Results

The results of the model in Table 1 show that the week, combined, and interaction effects all have a significant impact on forecasting error ($p < 0.05$). The negative coefficient indicates that learning might have indeed occurred: as time passes over the course of the semester, the average percentage forecasting error diminishes by 0.4% per week. This is equivalent to a 5.0% forecast error improvement over the 11-week period. Additionally, the expectation that an improvement in accuracy occurs due to the use of forecast combinations is supported as the error is improved by 4.5% when using a combination. The interesting result and focus of this investigation, however, relate to the positive coefficient of the interaction effect. It means that in the course of the semester,

Table 1. Learning-focused data: APE as a function of time, forecast combination, and Time × Combination.

Dependent variable APE		
Variable	Parameter	SE
Intercept	0.1547**	0.011
Combination	−0.0446***	0.013
Week	−0.0042***	0.001
Interaction (Combination × Week)	0.0033**	0.002
R^2	0.0184	
F-Test for no fixed effects	0.77	
Hausman test for random effects	0.29	
Unique student pairings	137	
Time series length	11	

Note: APE: average percentage error; SE: standard error.
**Significant at 0.05 level.
***Significant at 0.01 level.

the effect of the forecast combination, that is, the ability of the combination to improve accuracy, is reduced by 0.33%. For instance, the interaction effect on the last week decreases the accuracy by 3.96%. Taken together, the capacity of forecast combinations to improve accuracy when learning occurs diminishes from 4.1% (in week 1) to 0.5% (in week 12).

The model assumptions were reviewed through scatterplots, histograms, and residual plots, inferring that the model exhibited linearity, homogeneity of variance, while variance inflation factors indicated no concerns for multicollinearity. The normality of errors assumption appeared to be in question as the model lacked normality of the DV. Although skewness and kurtosis were in the designated range of positive and negative 2, the histogram of absolute percent error indicated that a transformation may improve estimates.

Given the concerns regarding normality, the DV of APE was reconstructed using a square root transformation. The necessary conditions for skewness and kurtosis were still upheld and the DV appeared (graphically) significantly more normal after the transformation. The regression model was refitted, and as Table 2 shows, the results are consistent with model 1 for the week and combined variables, with both measures significantly increasing forecasting accuracy ($p < 0.05$). Correspondingly, the interaction effect negatively affected forecasting accuracy with a positive coefficient that increased error; however, the p-value was only found significant at the 10% level ($p = 0.09$). In addition, the robustness checks of the model were all satisfied.

To summarize, hypothesis 1 appears to be partly supported by the findings of this preliminary study. In this unique environment where the forecasters are engaged in guided learning process with continuous feedback, learning seems to have occurred, and, as hypothesized, learning appears to reduce the effectiveness of forecasting combinations.

Study 2

Methodology

The second data set consisted of professional forecasting records of a 300-room Amsterdam (the Netherlands)-based hotel that recorded daily occupancy predictions over a period of 2 years. The

Table 2. Learning-focused data: Transformed APE as a function of time, forecast combination, and Time × Combination.

Dependent variable sqrt-APE		
Variable	Parameter	SE
Intercept	0.3596***	0.017
Combination	−0.0574***	0.02
Week	−0.0049**	0.002
Interaction (Combination × Week)	0.0042*	0.003
R^2	0.0113	
F-Test for no fixed effects	0.87	
Hausman test for random effects	0.43	
Unique student pairings	137	
Time series length	11	

Note: APE: average percentage error; SE: standard error.*Significant at 0.1 level.
**Significant at 0.05 level.
***Significant at 0.01 level.

hotel property recorded both human- (i.e. revenue manager) and machine (i.e. RMS software)-generated forecasts for multiple horizons over the sample period. Forecasting horizons ranged between 1 day and 97 days-ahead forecasts. These forecasting horizons of 1 to 97 days ahead were not uniformly distributed in the data set. The shorter forecasting horizons of 25 days or less appeared between 150 to 200 times per horizon, while the longer horizons (of over 25 days ahead) appeared less frequently with most of the horizons having less than 100 records. The lowest frequency was six records for each of the forecasting horizons of 95, 96, and 97 days ahead. In total, we utilized 2568 pairs of independent (human and machine) forecasts and the same number of combined forecasts (again using equal weights). Accordingly, the data set included a total of 7704 forecasts of daily hotel occupancy observations, and, of course, their corresponding actual numbers, that is, the actual occupancy figures. As with the novice's forecast competition data set, the DV used in the analysis of the hotel data was the APE. The model's independent variables included the following:

- a forecasting horizon—a variable ranging from 1day ahead to 97 days ahead;
- a combined forecast binary indicator—a categorical variable coded as 1 to denote forecast combination and 0 otherwise;
- the month the forecast was made—ranging from 1 to 25;
- an interaction term (month and combination status).

Results

Before fitting the full model, we tested whether learning occurred with the subjective predictions and with machine forecasts. The finding indicate that a small but statistically significant human learning effect occurred over time ($p < 0.001$) as the revenue managers' subjective predictions improved. Similarly, we find that the machine predictions improved over time ($p < 0.001$). The full model was fitted as an OLS regression to test the efficiency of using forecasting combinations

Table 3. Hotel professional data: APE as a function of forecast horizon, month, forecast combination, and Month × Combination.

Dependent variable APE		
Variable	Estimates	SE
Intercept	0.338***	0.007
Forecasting horizon	0.002***	0.000
Month	−0.013***	0.000
Combination	−0.105***	0.010
Interaction (Month × Combination)	0.006***	0.001
R^2	0.158	
Adjusted R^2	0.157	
N	7703	

Note: APE: average percentage error; SE: standard error.
***Significant at 0.01 level.

overtime. The DV, the forecast's APE, was modeled with the independent variables of forecasting horizon, combination indicator, month, and the interaction term. Table 3 reports the findings of the full model. It indicates that all of the tested variables have a statistically significant effect on forecasting accuracy ($p < 0.01$). The significance of the month variable suggests that learning and improvement over time may have occurred, with an improved forecasting accuracy (average reduction of the APE) of 1.3% per month. Over the course of the 25 studied months, the forecast accuracy in the last month increased by 32.5% compared to the first month. A simple linear combination of human-generated and computer-generated forecasts reduced the forecasting error (APE), that is, increased forecasting accuracy, by 10.5%. Additionally, the positive interaction term indicates that accuracy decreased by 0.6% per month. Thus, in the last month, the forecasting accuracy was decreased from the use of a combination by 15.0%. Combining this information with the previous accuracy improvements shows that the net accuracy in the last month, while using combinations, is actually a net accuracy decrease of 4.5%. The capacity of forecast combinations to improve accuracy when learning occurs thus diminishes from a decrease in forecast error of −9.9% in month 1, a decrease that gradually becomes less with each subsequent month, to an actual increase in forecast error of 4.5% in month 25. As such, the positive impact of learning is not 32.5% but 28% in that last month. Additionally, the model reveals a small but statistically significant coefficient of 0.002 for the forecasting horizon, agreeing with previous published research that forecasts tend to lose accuracy with larger forecasting horizons.

When diagnosing the robustness of the model, the residual plots confirmed many of the assumptions for regression; however, the q–q plot, histograms, skewness, and kurtosis metrics suggested that the normality of error assumption may be violated. In an attempt to correct the assumption, a two-step approach to transforming the absolute percent error term was taken to obtain the best possible model. Since the error terms ranged from 0% to 275%, transformations had different impacts on values above and below 1. Accordingly, extreme errors were winsorized back to 1. This accounted for 48 observations or less than 1% of the data. Additionally, the values needed to be transformed to obtain normalized characteristics. A square root transformation was chosen, which provided skewness and kurtosis values within the positive and negative 2 range and a considerably more acceptable normality plot.

Table 4. Hotel professional data: Transformed APE as a function of forecast horizon, month, forecast combination, and Month × Combination.

Dependent variable sqrt-APE		
Variable	Estimates	SE
Intercept	0.509***	0.007
Forecasting Horizon	0.002***	0.000
Combination	−0.06***	0.010
Month	−0.012***	0.000
Interaction (Month × Combination)	0.003***	0.001
R^2	0.159	
Adjusted R^2	0.159	
N	7703	

Note: APE: average percentage error; SE: standard error.
***Significant at 0.01 level.

A second regression model was fitted with the transformed DV using the same independent variables. The results (see Table 4) are consistent with the prenormalized variables in terms of direction and significance. The variables month and the combination indicator both significantly reduced forecasting error within the model ($p < 0.01$). The interaction term also had a significant positive effect, suggesting that forecasting error increased overtime from combination ($p < 0.01$). Finally, the horizon variable remained unchanged from the previous model ($p < 0.01$).

Similar to the results of study 1, the findings of study 2 appear to provide initial support for the hypothesis that learning reduces the effectiveness of forecast combinations in producing more accurate predictions.

Study 3

Methodology

The data for the main study were provided by a globally recognized RMS vendor. It contained machine-generated forecasts at predefined horizons and user overrides for some, but not all, of the system's forecasts. More specifically, the fields contained property, segment, and user identifiers for overrides, the day of occupancy, the day of forecast, the day of an override and the type of override, as well as the actual number of rooms sold. For the analysis, we defined, and calculated, the following:

Forecasting horizon is the difference (in days) between the day the system forecast (or the user override) was made and the date of stay. Forecasting horizon is being controlled for because the literature suggests that the longer the horizon the less accurate the forecast (e.g. Döhrn and Schmidt, 2010; Smith and Sincich, 1991).

User learning was assessed using a counter that captured each successive "override day."

Machine learning was assessed by a counter that tracked each successive forecast for a property at a given forecasting horizon.

Moreover, the data set was split into two subsets: system only forecasts and user override data.

The system forecast data (non-override) was analyzed ahead of the overrides data to assess whether machine learning occurs, that is, if the accuracy of the computer-generated forecasts improved over time. This subset of data had over 20 million (20,003,446) initial observations. Each day of occupancy had several forecasts conducted at fixed horizons. Seven different horizons (denoted horizon 1–7) are analyzed ranging from zero days, that is, a forecast on the day of stay, to a maximum of eight months in advance. After filtering out nonusable records such as records with no overrides, no actual stays, and forecast of zero, less than eight million (7,886,262) observations remained.

The overrides subset consisted of 120,106 overrides from 1274 hotels generated in the period of November 2013 to May 2016. To measure the impact of learning by the user, each property ID was paired with a corresponding user ID, generating 2611 unique person/property pairings across the 1274 locations. Several overrides must occur over time to allow for feedback and learning to occur. Accordingly, individuals who conducted less than four overrides (roughly 50% of the override sample) were removed from the analysis. In total, the final data set consisted of 91,759 observations from 900 hotels with 694 experts, generating 1321 unique person/property pairings, and overrides per person ranging from 4 to 286.

Dependent variable. The accuracy of the system, that is, of machine generated forecasts, was measured as follows: $\log(\text{system APE} + 1)$. The efficiency of forecast overrides was measured as the change in error percentage between the RMS and the user. Accordingly, the DV was defined as follows:

$$\text{DV} = \log(\text{system APE} + 1) - \ln(\text{user APE} + 1)$$

Note that 1 was added to the transformation of the error so that a "perfect" forecast would correspond to $\log(1) = 0$, avoiding the asymptotic predicament of taking the natural log of 0. Since the user's error metric is subtracted from the system's error metric, a positive value for the DV means that the user override improved the accuracy of the machine forecast.

Independent variables. Several independent variables were used in the two models, including the forecasting horizon, the forecast counter, and the number of overrides, as well as an interaction effect. The forecast counter and the number of overrides captured the learning of the machine and user, respectively.

Models. To asses whether learning occurred with the system forecasts, the following model (1) was fitted separately for each of the seven forecast horizons, where a negative coefficient of the sequential forecast counter indicates that the software learns, that is, the more forecasts generated and evaluated the more accurate the forecasts:

$$\log(\text{system APE} + 1) = a + \text{forecast counter} + \varepsilon \tag{1}$$

To assess the main research question of this study, that is, to find whether learning decreases the ability of forecast combinations (i.e. user overrides) to improve accuracy, the following model (2) was fitted:

$$\begin{aligned} \text{Difference in absolute percent log error[DV]} = a &+ \text{long horizon} + \text{override day} \\ &+ \text{Long Horizon} \times \text{Override Day} \end{aligned} \tag{2}$$

The new IV, "long horizon," was a collapsed, binary, version of the "override horizon" variable, where overrides with horizons less than 21 days each had a value of 0 and horizons greater than 21

Table 5. System accuracy over time.

Variable	Horizon 1	Horizon 2	Horizon 3	Horizon 4	Horizon 5	Horizon 6	Horizon 7
Intercept	0.0437	0.0448	0.0718	0.1230	0.2214	0.3462	0.3850
Forecast counter	−0.002***	−0.002***	−0.002***	−0.002***	0.002***	0.0102***	0.0115***
R^2	0.187	0.158	0.118	0.106	0.146	0.180	0.200
F-Test for no fixed effects	135.76***	105.46***	85.17***	80.71***	123.16***	162.98***	182.59***
Hausman test	1259.62***	436.44***	281.03***	205.36***	410.94***	587.63***	725.47***
Locations	1512	1487	1489	1489	1487	1487	1489
Time series length	1106	1092	1102	1110	1127	1139	1140
N	960,854	872,175	967,925	1,030,462	1,101,710	1,139,729	1,128,216

Note: Dependent variable = log (system APE + 1).
**Significant at 0.05 level.
***significant at 0.01 level.

days had the value of 1. This horizon grouping was based on the findings of model 1, as will be explained in the Results section.

The cross section of data over time required a panel model for adequate specification. The data were considered unbalanced as override frequency varied across users. Several tests were used to determine whether a fixed effects or random effects model was more appropriate. Most notably, the Hausman test for random effects was significant ($p < 0.05$), indicating that the fixed effects model should be used. This outcome allows the model to control for all the unique characteristics of the properties while still being able to capture the overall effect of override efficiency over time.

Results

The findings of model 1 (See Table 5) are mixed: shorter forecasting horizons, that is, horizons 1–4 (21 or less days ahead), have significant reductions in system error, while longer forecasting horizons, that is, horizons 5–7, appear to have increases in forecast error. In other words, the findings appear to suggest that machine learning improves forecasting of shorter horizons, but there is no evidence to suggest that it happens in the longer forecasting horizons.

The final model (2) tests the hypothesis that learning diminishes the capacity of user overrides to improve (system) forecast accuracy, and the results are presented in Table 6. Recall that given [DV], a positive value indicates that the user override improved the accuracy, while a negative one indicates the machine forecast was more accurate than the subjective adjustment.

The assessed impact of the control variable of horizon is as follows. The negative effect of the long horizon indicates that user overrides better correct the machine predictions in short horizons, compared to long horizons greater than 3 weeks. This indicate that overrides done earlier are less likely to correct the machine predictions compared to overrides closer to the date of stay.

The override day coefficient is statistically insignificant, indicating that user learning does not occur, while the interaction element of "Override Day × Long Horizon" is positive at the $p < 0.01$ level. Taken together, the interpretation of these two is as follows. It appears that learning does not significantly impact the effectiveness of overrides in the short-term horizons. However, compared

Table 6. Override efficiency fixed effects model.

Variable	Parameter	SE
Intercept	0.3189	0.1581
Long horizon $\geq$ 21 days (horizons 5, 6, 7)	−0.1381***	0.0103
Override day	0.0001	0.0002
Interaction Long Horizon $\times$ Override Day	0.0007***	0.0002
R^2	0.0444	
F-Test for no fixed effects	3.02***	
Hausman test for random effects	19.09**	
Locations	1321	
Time series length	1480	

Note: Dependent variable (forecast error improvement) = ln (system APE + 1) − ln (user APE + 1). SE: standard error.
**Significant at 0.05 level.
***Significant at 0.01 level.

to the short term, in the long-term horizons of over 21 days in advance, learning increases the effectiveness of forecast combination, that is, the likelihood of user overrides to improve forecast accuracy. In other words, with RMS overrides (as opposed to the more traditional combination scenarios tested in the first two preliminary studies) the hypothesis is not supported. Learning appears to have no effect on combination effectiveness in shorter horizons, and in the longer forecasting horizons, it increases, rather than reduces, the effectiveness of forecast combinations.

Discussion

The results of the analyses present several important findings for the field of RM. First consider the finding of study 3 that machine-algorithmic learning and the consequent improvement in forecast accuracy appear to happen in the short forecasting horizons but not in the longer ones. It is not easy to explain these contradicting results. In fact, intuition might suggest the opposite: In longer horizons where forecasting is more challenging, repetition and feedback-based learning should have resulted in a detectable improvement in forecast accuracy more (and not less) than in the shorter forecasting horizons. We guess that this surprising result might have to do with the overall difficulty associated with long forecasting horizons, that is, the notion that overall, forecasts of longer horizons are less accurate. It could be that similarly, RMSs may be more adept to incorporate updated information over time, that is, to learn, at shorter horizons. Alternatively, deterioration of accuracy, when learning and improvement is supposed to occur, might be due to shifts in the booking window. By definition, the longer the forecasting horizon, the more likely it is that the booking window has changed during that period. These booking window shifts have been suggested (Schwartz et al., 2016; Webb, 2016) as a potential source of “confusion” for forecasting models because, by the nature of the change, “misleading” information is introduced.

The results of the three studies are mixed in that the support for the hypothesis appears to be inconsistent. The first two preliminary studies support the notion that as learning occurs, the effectiveness of forecast combination in generating more accurate forecasts is reduced. Conversely, the results of the main study suggests that in short-term horizons (21 days or less) learning has no impact on the effectiveness of forecast combinations, while in longer forecasting horizons the impact of learning is positive—the more learning, the more effective the forecast combinations in generating accurate forecasts. We suggest that the difference might be due to the nature of the combinations. In the first two scenarios, the combinations adhere to the idle condition of the elements of the forecast combination are independent of each other. The student teams generating the occupancy forecasts have not shared information, did not necessarily apply the same forecasting methods, and might have had nonshared private information. Similarly, in the second study, the RM team members who generated subjective forecasts during the study's period did not observe the computer-generated forecasts, while the computer algorithm was made aware of the human subjective assessments. However, with the more realistic, and common, scenario tested in study 3, the assumption of independence is violated. By definition, user override of algorithmic prediction can't be independent of the machine forecast. Accordingly, when considered together, the results of the two tests of study 3 might suggest the following about the longer forecasting horizons: since there is no machine learning in the longer horizons (the results of the first model of study 3), if users do learn over time, the overrides (combinations) effectiveness should indeed be improving. In the short-term horizons, where according to the first model of study 3, machine learning occurs, it does make sense that no change in override effectiveness is observed as time goes by when the system is inherently improving on its own.

From a practical perspective, the implications for hotel practices are as follows. When considering the more traditional (not the overrides) types of forecast combinations, where the elements are independent, novice forecasters, as is often the case with hotels that are just starting to implement systematic RM practices, are more likely to benefit from forecasting combinations compared to hotels that have been at it for some time and were consciously, and systematically, monitoring the accuracy of their forecast, striving to improve by employing various approaches. Similarly, and as important, a revenue team in a "new" hotel environment (hotel recently opened, acquired, or switched management company) may also benefit considerably from combining forecasts. Second, we already know that hotel RM teams, and RM solution providers, should frequently monitor the accuracy of their forecasts. What we learned from this study is that the assumption that forecast combinations are always likely to enhance accuracy is probably erroneous. Given these observed relations, and the notion that in reality different kinds and levels of forecasting learning might occur, it follows that hotels need to monitor not only the overall forecast accuracy performance of the RM system but also look closely, and frequently, at the combination's contribution to accuracy. Our findings indicate that as time goes by, and if learning and training occurs, that combination contribution is likely to diminish. In some cases, the hotel might be better off not combining, underscoring the importance of monitoring the effectiveness of forecast combinations. If learning and training does not appear to affect the efficacy of forecast combinations, the hotel might conclude that learning and training is inefficient and should therefore be improved. Alternatively, it could mean that the reason for the combination's contribution to accuracy is not altered by learning, for example, the private information does not become more public with training.

When the combination includes the user override, it is clear that learning appears to be important and more beneficial when forecasting at longer horizons. Since these horizons are the more challenging to forecast to begin with, it is important for the RM team not only to monitor the accuracy in these horizons but also to train the revenue managers and provide frequent feedback to generate learning as the overrides are likely to be more effective in these longer horizons.

Limitations and future research

This study's original contribution spans several areas. We are first to study the effectiveness of forecast combinations when learning occurs, discussing the theoretical underpinning and conducting empirical tests. In addition, we are first to demonstrate that the common practice of overrides of algorithmic predictions (the common way in which forecast combinations are practiced in the hotel industry) might be different from the traditional forecast combination scenarios where the combined elements are independent, or semi-independent, of each other. As this is a first attempt to explore the learning/combination interaction and look at scenarios where the elements of the combination are not independent of each other, there are various issues and questions that were not addressed and should be explored in the future.

Repetition was a proxy for learning in two of the studies. That is, we have no direct observation and verification of how machine and humans learned. While the data providers reported that the forecasting algorithms used have a learning element, we are not able to verify that aspect since they would not disclose relevant details. We assumed that revenue managers and machines are learning by generating more forecasts and by receiving feedback on the accuracy. Furthermore, we have verified that accuracy improvement overtime occurred. This is a strong indication that learning indeed occurred. However, this aspect of learning should be studied more deeply in the future because different learning processes might have a different impact on the effectiveness of the forecast combination. In this regard, the use of an experimental design should be considered.

The forecast's time horizon was modeled as a categorical variable. That is, we binned different horizons, in part, because of the nature of the data provided by the software company and to maintain confidentiality. Future research with perhaps different data could look more granularly at the forecast horizon variable and generate more specific insights on the learning impact on the effectiveness of forecast combination in different forecast horizons.

Finally, what we learn in this study points to the importance of conducting more research on many aspects of user overrides of machine forecasts and recommendations in hotel RM. To the best of our knowledge, very little is published about this important topic.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Zvi Schwartz https://orcid.org/0000-0001-9254-5670
Timothy Webb https://orcid.org/0000-0002-4435-1596

References

Aiolfi M, Capistran C and Timmermann A (2011) Forecast combinations. In: Clements P and Hendry D (eds) *The Oxford Handbook of Economic Forecasting*. Oxford: Oxford University Press, pp. 355–390.

Armstrong JS (1986) The ombudsman: research on forecasting: a quarter-century review, 1960-1984. *Interfaces* 16(1): 89–103.

Armstrong JS (2001a) Combining forecasts. In: Armstrong SJ (eds) *Principles of forecasting: a handbook for researchers and practitioners*. Dordrecht: Kluwer Academic, pp. 417–440.

Armstrong JS (2001b) Standards and practices for forecasting. In: Armstrong SJ (eds) *Principles of forecasting: a handbook for researchers and practitioners*. Dordrecht: Kluwer Academic, pp. 679–732.

Arora H (2018) 7 tips to increase revenue at your hotel. Available at: https://letsrockthehotel.com/7-tips-to-increase-your-hotel-revenue/ (accessed 6 July 2019).

Bates JM and Granger CWJ (1969) The combination of forecasts. *Operational Research Quarterly* 20(4): 451–468.

Bendoly E (2013) Real-time feedback and booking behavior in the hospitality industry: moderating the balance between imperfect judgment and imperfect prescription. *Journal of Operations Management* 31(1/2): 62–71.

Boweden R (1989) Feedback forecasting games: an overview. *Journal of Forecasting* 8(2): 117–127.

Chu FL (1998) Forecasting tourism: a combined approach. *Tourism Management* 19: 515–520.

Christoffersen P and Diebold FX (1997) Optimal prediction under asymmetrical loss. *Econometric Theory* 13: 806–817.

Clemen RT (1987) Combining overlapping information. *Management Science* 33(3): 373–380.

Clemen RT (1989).Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting* 5(4): 559–583.

Collopy F and Armstrong JS (1992) Rule-based forecasting: development and validation of an expert systems approach to combining time series extrapolations. *Management Science* 38(10): 1394–1414.

Collopy F, Armstrong JS and Adya M (2001) Expert systems for forecasting. In: Armstrong SJ (eds) *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Dordrecht: Kluwer Academic, pp. 283–300.

Cross R, Higbie J and Cross D (2009) Revenue management's renaissance: a rebirth of the art and science of profitable revenue generation. *Cornell Hospitality Quarterly* 50(1): 56–81.

Dacko SG (2004) Marketing strategies for last-minute travel and tourism: profitability and revenue management implications. *Journal of Travel & Tourism Marketing* 16(4): 7–20.

Diebold FX and Lopez JA (1996) Forecast evaluation and combination. In: Maddala GS and Rao CR (eds) *Statistical Methods in Finance, Handbook of Statistics Vol 14*. Amsterdam: Elsevier, pp. 241–268.

Döhrn R and Schmidt CM (2010) Information or institution? – On the determinants of forecast accuracy. *Ruhr Economic Papers #201*. Available at: https://www.econstor.eu/bitstream/10419/45323/1/635642387.pdf (accessed 10 February 2019).

Fairly T (2013) Take it further. Total hotel revenue management. Available at: http://www.hsmai.org/knowledge/summary.cfm?ItemNumber=10367 (accessed 10 January 2015).

Fildes R, Nikolopoulos K, Crone SF, et al. (2008) Forecasting and operational research: a review. *Journal of the Operational Research Society* 59(9): 1150–1172.

Fischhoff B (2001) Learning from experience: coping with hindsight bias and ambiguity. In: Armstrong SJ (eds) *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Dordrecht: Kluwer Academic, pp. 543–544.

Forgacs G (2010) *Revenue Management: Maximizing Revenue in Hospitality Operations*. Lansing: The American Hotel & Lodging Institute.

Franses PH (2014) *Expert Adjustments of Model Forecasts: Theory, Practice and Strategies for Improvement*. Cambridge: Cambridge University Press.

Genre V, Kenny G, Meyler A, et al (2013) Combining expert forecasts: can anything beat the simple average? *International Journal of Forecasting* 29(1): 108–121.

Goodwin P (2009) New evidence on the value of combining forecasts. *Foresight: The International Journal of Applied Forecasting* 12: 33–35.

Granger CWJ and Ramanathan R (1984) Improved methods of combining forecasting. *Journal of Forecasting* 3(2): 97–204.

Haensel A and Koole G (2011) Booking horizon forecasting with dynamic updating: a case study of hotel reservation data. *International Journal of Forecasting* 27(3): 942–960.

Hibon M and Evgeniou T (2005) To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting* 21(1): 15–24.

Hoeting JA, Madigan DM, Raftery AE, et al. (1999) Bayesian model averaging: a tutorial. *Statistical Science* 14(4): 382–417.

Koupriouchina L, van der Rest JI and Schwartz Z (2014) On revenue management and the use of occupancy forecasting error measures. *International Journal of Hospitality Management* 41: 104–114.
Lahiri K, Peng H and Zhao Y (2013) *Machine learning and forecast combination in incomplete panels*. Available at: SSRN:https://ssrn.com/abstract=2359523 (accessed 10 September 2019).
Law R (1998) Room occupancy rate forecasting: a neural network approach. *International Journal Of Contemporary Hospitality Management* 10(6): 234–240.
Law R (2000) Demand for hotel spending by visitors to Hong Kong: a study of various forecasting techniques. *Journal of Hospitality and Leisure Marketing* 6(4): 17–29.
Makridakis S (1989) Why combining work? *International Journal of Forecasting* 5: 601–603.
Makridakis S and Hibon M (2000) The M3-competition: results, conclusions and implications. *International Journal of Forecasting* 16(4): 451–476.
Mayock P (2009) 10 hotel booking trends and other travel market developments. Available at: http://www.hotelnewsnow.com/Articles.aspx/1673/10-hotel-booking-trends-and-other-travel-market-developments (accessed 12 January 2012).
Mayock P (2011) Sales teams feel pinch of short booking windows. Available at: http://www.hotelnewsnow.com/articles.aspx/5932/Sales-teams-feel-pinch-of-short-booking-windows (accessed 12 January 2012).
Mukhopadhyay S, Samaddar S and Colville G (2007) Improving revenue management decision making for airlines by evaluating analyst-adjusted passenger demand forecasts. *Decision Sciences* 38(2): 309–327.
Newbold P and Harvey DI (2002) Forecast combination and encompassing. In: Clemenets MP and Hendry DF (eds) *A companion to Economic Forecasting*. Oxford: Blackwell, pp. 268–283.
Padhi SS and Aggarwal V (2011) Competitive revenue management for fixing quota and price of hotel commodities under uncertainty. *International Journal of Hospitality Management* 30(3): 725–734.
Pereira LN (2016) An introduction to helpful forecasting methods for hotel revenue management. *International Journal of Hospitality Management* 58(3): 13–23.
Prudêncio R and Ludermir T (2005) Using machine learning techniques to combine forecasting methods. In: Webb GI and Yu X (eds) *AI 2004: Advances in Artificial Intelligence Berlin: Springer*, pp. 1122–1227.
Remus W, O'Conner M and Griggs K (1996) Does feedback improve the accuracy of recurrent judgmental forecasts? *Organizational Behavior and Human Decision Processes* 66(1): 22–30.
Revenue Analytics (2018) Keeping overrides from becoming overkills for revenue management systems. Available at: https://www.revenueanalytics.com/ra-blog/keeping-overrides-becoming-overkills-revenue-management (accessed 6 February 2019).
Rieg R (2008) Measuring improvement in forecast accuracy: a case study. *Foresight: The International Journal of Applied Forecasting* 11: 15–20.
Schwartz Z (2003) Subjective estimates of occupancy forecast uncertainty by hotel revenue managers. *Journal of Travel & Tourism Marketing* 16(4): 59–66.
Schwartz Z and Cohen E (2004) Hotel revenue-management forecasting. *Cornell Hotel & Restaurant Administration Quarterly* 45(1): 85–98.
Schwartz Z, Uysal M, Webb T, et al. (2016) Hotel daily occupancy forecasting with competitive sets: a recursive algorithm. *International Journal of Contemporary Hospitality Management* 28(2): 267–285.
Shen S, Li G and Song H (2008) An assessment of combining tourism demand forecasts over different time horizons. *Journal of Travel Research* 47(2): 197–207.
Shen S, Li G and Song H (2010) Combination forecasts of international tourism demand. *Annals of Tourism Research* 38(1): 72–89.
Smith SK and Sincich T (1991) An empirical analysis of the effect of length of forecast horizon on population forecast errors. *Demography* 28(2): 261–274.
Smith J and Wallis KF (2009) A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics* 71(3): 331–355.
Song H, Witt SF, Wong KF, et al. (2009) An empirical study of tourism demand forecast combination. *Journal of Hospitality and Tourism Research* 33(1): 3–29.

Stewart T (2001) Improving reliability of judgmental forecasting. In: Armstrong SJ (eds) *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Dordrecht: Kluwer Academic, pp. 81–106.

Stock J and Watson MA (1999) Comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In: Engle R and White H (eds) *Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger*. Oxford: Oxford University Press, pp. 1–44.

Stock JH and Watson M (2004) Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* 23(6): 405–430.

Talluri KT and van Ryzin GJ (2006) *The Theory and Practice of Revenue Management*. Springer Science & Business Media.

Timmermann A (2006) Forecast combinations. In: Elliott G, Granger CWJ and Timmermann A (eds) *Handbook of Economic Forecasting*. Amsterdam: Elsevier, pp. 135–196.

Weatherford LR and Kimes SE (2003) A comparison of forecasting methods for hotel revenue management. *International Journal of Forecasting* 19(3): 401–415.

Webb T (2016) From travel agents to OTAs: how the evolution of consumer booking behavior has affected revenue management. *Journal of Revenue and Pricing Management* 15(3-4): 276–282.

Wong KF, Song H, Witt SF, et al. (2007) Tourism forecasting: to combine or not to combine? *Tourism Management* 28(4): 1068–1078.

Yang Y (2004) Combining forecasting procedures: some theoretical results. *Econometric Theory* 20(1): 176–222.

Zellner A (1986) Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association* 81(394): 446–451.

Zheng T, Bloom BN, Wang X, et al. (2012) How do less advanced forecasting methods perform on weekly RevPAR in different forecasting horizons following the recession? *Tourism Analysis* 17(4): 459–472.

Author biographies

Zvi Schwartz is a professor in the Department of Hospitality Business Management at the University of Delaware. Prior to joining the University of Delaware, Dr Schwartz was the J. Willard and Alice S. Marriott Senior Faculty Fellow for Hospitality Finance and Revenue Management in the department of Hospitality and Tourism Management at Virginia Tech where he served as the director of graduate programs and an associate professor at the University of Illinois. Prior to joining academia, he had over a decade of lodging industry experience as a manager and an entrepreneur.

Timothy Webb is an assistant professor of Hospitality Business Management at The University of Delaware. His research is focused on the fields of revenue management and analytics in the hospitality industry. He recently graduated with his PhD in Hospitality and Tourism Management from Virginia Tech and was formerly employed as a Data Scientist for Delaware North, a global hospitality company.

Jean-Pierre I van der Rest is a professor of Business Administration at Leiden University. He holds a concurrent position as Dean at the Graduate School and previously served as Professor, Research Director, and Associate Dean at Hotelschool The Hague. He serves on various editorial boards and received research awards for his work on pricing decision-making and revenue optimization in the hotel and restaurant industry. He has taught in Bachelor, Master, MBA, and Executive Education programs, supervised and examined doctoral theses, has chaired accreditations, has been an external examiner, external advisor, and a visiting professor at various international business schools.

Larissa Koupriouchina is a senior lecturer in Revenue Management and Research fellow at Hotelschool The Hague. Next to teaching, she is working on her PhD degree at Leiden University, the Netherlands. Her doctoral work is sponsored by the research grant of the Netherlands Organization for Scientific Research and focuses on human–computer interactions between revenue managers and revenue management systems. Before starting her career in education, she worked for over 10 years on revenue and profit recovery projects.