BOOK OF SHORT PAPERS

3rd ITALIAN CONFERENCE ON ECONOMIC STATISTICS

SUSTAINABILITY, INNOVATION AND DIGITALIZATION:
Statistical Measurement for Economic Analysis

UNIVERSITÀ DEGLI STUDI DI NAPOLI PARTHENOPE

DISAQ DIPARTIMENTO DI STUDI AZIENDALI E QUANTITATIVI DIPARTIMENTO DI ECCELLENZA

SIS Società Italiana di Statistica

ENZO ALBANO EDIZIONI

ICES 2025 - 3rd Italian Conference on Economic Statistics

BOOK OF SHORT PAPERS

Editors: Rosalia Castellano, Giovanni De Luca, Emma Bruno

Book of Short Papers
3rd Italian Conference on Economic Statistics (**ICES 2025**)
Sustainability, Innovation and Digitalization:
Statistical Measurement for Economic Analysis

UNIVERSITÀ DEGLI STUDI DI NAPOLI PARTHENOPE

DISAQ DIPARTIMENTO DI STUDI AZIENDALI E QUANTITATIVI
DIPARTIMENTO DI ECCELLENZA

SIS Società Italiana di Statistica

ENZO ALBANO
EDIZIONI

Organization

Scientific and Program Committee

Rosalia Castellano (Chair)	University of Naples Parthenope
Matilde Bini	European University of Rome
Vincenzo Candila	University of Salerno
Marcello Chiodi	University of Palermo
Andrea Ciccarelli	University of Teramo
Corrado Crocetta	University of Bari
Giovanni De Luca	University of Naples Parthenope
Elena Fabrizi	University of Teramo
Francesca Giambona	University of Florence
Chiara Gigliarano	Carlo Cattaneo University
Caterina Marini	University of Bari
Andrea Marletta	University of Milano-Bicocca
Vittorio Nicolardi	University of Bari
Paolo Mariani	University of Milano-Bicocca
Antonella Rocca	University of Naples Parthenope
Giuseppe Storti	University of Salerno
Giorgio Tassinari	University of Bologna
Roberto Zelli	Sapienza University of Rome

Local Organizing Committee

Giovanni De Luca (Chair)	University of Naples Parthenope
Rosalia Castellano	University of Naples Parthenope
Sergio Longobardi	University of Naples Parthenope
Paolo Mazzocchi	University of Naples Parthenope
Margherita Pagliuca	University of Naples Parthenope
Gennaro Punzo	University of Naples Parthenope
Andrea Regoli	University of Naples Parthenope
Giorgia Rivieccio	University of Naples Parthenope
Antonella Rocca	University of Naples Parthenope
Giuseppe Scandurra	University of Naples Parthenope
Emma Bruno	University of Naples Parthenope
Anna Pia Di Iorio	University of Naples Parthenope
Michele Mario Ippolito	University of Naples Parthenope
Andrea Montanino	University of Naples Parthenope
Giuseppe Verdoliva	University of Naples Parthenope

Leveraging Data Integration for Effective Revenue Management

Andrea Guizzardi, Ercolino Ranieri and Federico Zardo

Abstract In this study, we seamlessly integrate micro-level data from both internal and external sources to revolutionize last-minute occupancy forecasting across the advance booking horizon. By utilizing a sophisticated beta regression framework coupled with a robust variable selection methodology, we highlight the importance of feature engineering raw data. This approach delivers actionable insights that drive revenue optimization and elevate performance in the hospitality sector.

Key words: Hospitality sector, Occupancy forecasting, Advance bookings, Feature engineering, Beta regression

1 Introduction

Forecasting demand in the hotel industry poses a significant challenge for revenue managers due to irregular fluctuations driven by promotions, events, and environmental changes, which affect sales processes. This study highlights the importance of integrating Property Management Systems (PMS) with external sources of information to increase the accuracy of occupancy prediction models. We acknowledge that for many small and medium-sized hotels, system integration requires a mindset shift that is often difficult to implement due to limited financial, technical, and human capital resources. However, this is a necessary change for effective pricing

Andrea Guizzardi
University of Bologna, e-mail: andrea.guizzardi@unibo.it

Ercolino Ranieri
Xenia Spa, e-mail: ercolino.ranieri@xeniahs.com

Federico Zardo
Xenia Spa, e-mail: federico.zardo@xeniahs.com

and resource allocation decisions, in light of the growing use of digital tools by consumers in choosing their destination and accommodation [3].

By examining a medium-sized, four-star hotel in central Italy that offers two specific room types—*classic* and *superior*—we demonstrate that reliable forecasts can be achieved through accurate feature engineering of raw data using relatively simple methods like beta regression, rather than complex statistical methodologies that are harder to implement for many small and medium-sized hotels.

2 Data Description and Methodology

In this work, we suggest a feature engineering approach to the specification of a new type of independent variables with the idea that this could significantly influence the accuracy in predicting the daily pick-up curve. We consider daily data spanning from August the 12th 2022, marking the end of the COVID period, to December the 3rd 2024. Our dependent variable ($O_{t,k}$) represents the proportion of occupied rooms to the total available rooms on the arrival day t, with advance booking $k = 0$. We forecast $O_{t,0}$ at two horizons, $k = 5$ and $k = 10$ days ahead, a critical timeframe for decision-making due to high last-minute volatility [2]. As independent variables, we consider a combination of micro- and macro-level data, providing a holistic perspective that accounts for both internal and external determinants of tourism demand. Micro-level variables ($O_{t,k}$ and pricing strategies $P_{t,k}$) capture the temporal dynamics influenced by management choices, while macro-level data, such as monthly domestic and inbound tourism demand at the destination, reflect the destination's market size.

Micro-level data are synthesized using averages and standard deviations, across four predetermined booking intervals: $[5,10),[10,19),[19,38),[38,345)$. This aggregation approach is justified by the consistent dynamics observed between adjacent booking days. Notably, statistics from the $[5,10)$ interval are excluded for the 10-day ahead forecasts. We also include average cancellation rates calculated over longer intervals: $[5,345)$ for the 5-day forecast and $[10,345)$ for the 10-day forecast.

Borrowing common practices from financial markets, we perform feature engineering proposing technical-analysis indicators. Moving averages are particularly valuable for enhancing the revenue management practice of using the same day's performance from the previous year as a benchmark for current performance. Specifically, we define the moving average as $\mathrm{MA}(y_t) = \frac{1}{2k+1}\sum_{j=-k}^{k} y_{t-j-365}$, where $k = 5$ corresponds to a 10-day window. This formulation allows us to better capture seasonality in prices and occupancy rates by smoothing short-term fluctuations. We also capture the variables 'momentum' with the following Stochastic-like and MACD-like oscillators: $\mathrm{Stoch}(X) = \frac{X_{t,0}-X_{t,\min}}{X_{t,\max}-X_{t,\min}}, \quad \mathrm{MACD}(Z) = \frac{Z_{[a,b]}}{Z_{[a,c]}}$, where $X_{t,0}$ is the most recently observed price or occupancy, $X_{t,\min}$ and $X_{t,\max}$ are the minimum and the maximum observed values, respectively. Here, $Z_{[i,j]}$ may denote the mean or the standard deviations of both ($O_{t,k}$ or $P_{t,k}$) in the advance booking interval [i,j) (e.g., $[5,10)$). These indicators are computed for both 'classic' and 'complementary' room types. After this preprocessing phase, we obtained 47 and 52 explanatory variables for the forecasts at horizons $k = 5$ and $k = 10$, respectively.

In our Beta regression approach $y_t \sim \text{Beta}(\mu_t, \phi)$,, $\mu_t \in (0,1)$ represents the mean and $\phi > 0$ denotes the precision. μ_t is linked to explanatory variables X_t via a logit link function. $\mu_t = \frac{\exp(X_t\beta)}{1+\exp(X_t\beta)}$.

Model parameters are estimated using maximum likelihood [1]. Significant variables are selected by fitting the model across 100 random partitions of the dataset (80% training, 20% testing). For each partition, a stepwise selection procedure guided by the AIC criterion identifies significant variables. The frequency of each variable's inclusion across the 100 regressions ranks their importance. For the final model, the top 20 most important variables are selected, a choice that does not significantly impact our results. Forecast accuracy is assessed using a distribution representing the Mean Absolute Error (MAE) from 1000 random dataset partitions (80% training, 20% testing), with occupancy rates converted back to the number of rooms for interpretability.

3 Results and conclusion

We report aggregate results for both *Classic* (32 rooms) and *Superior* (25 rooms), totaling 57 rooms. The histograms below represent 1,000 MAEs associated with each partition of the dataset. They clearly show that accuracy decreases as we move further from the check-in date, influenced by late cancellations and last-minute bookings. Considering the high market volatility in the last minute, the performance appears satisfactory, with a maximum error of less than 5% of the available rooms for the 10-day forecasts.

	Min	1st Q.	Median	Mean	3rd Q.	Max
MAE	1.627	1.903	2.109	2.109	2.194	2.409

	Min	1st Q.	Median	Mean	3rd Q.	Max
MAE	2.193	2.437	2.787	2.787	2.896	2.844

Fig. 1. MAE distributions: 5-day Forecast (left) - 10-day Forecast (right)

Table 1 provides a structural overview of the results, grouping the 20 variables that are most frequently significant in the 1,000 estimates into macro-categories. Notably, occupancy-related variables are the most informative, with 9 out of them in the ‘top 20’ significant variables at the 5-day horizon and 6 at the 10-day horizon. Occupancy-related metrics also show high relevance: on average, they appear significant in 79% of the estimated models for the 5-day horizon and in 84% for the 10-day horizon.

Table 1 Structural determinants

	5-day Forecast		10-day Forecast	
By macro categories	N°	(%)	N°	(%)
Occupancy	9	79%	6	84%
Prices	6	57%	7	57%
Context	2	60%	4	51%
Moving Avarages	1	100%	1	89%
Oscillators	2	52%	2	50%

Price-related indicators and contextual factors are influential across both forecast horizons (although they are fewer). The moving average measure achieves an impressive 100% score on the 5-day forecasting horizon, highlighting the relevance of our ‘fuzzy’ measure of the previous year’s occupancy.

Overall, Table 1 shows that integrating macro and micro data sources enhances forecast accuracy. However, the specified data generation process is mostly autoregressive, as occupancy metrics (both recent and those referring to the previous year) play a central role in capturing booking dynamics. This autoregressive component became stronger in the 5-day ahead forecasts, while 10-day ahead forecasts rely more on broader trends and macro-level factors. From a managerial point of view, this result suggests that real-time pricing adjustments are more effective in the short term, offering actionable insights for implementing techniques to refine pricing and inventory decisions.

References

1. Ferrari, S., Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799–815.
2. Guizzardi, A., Angelini, G., Pons, F.M.E. (2020). Does advance booking matter in hedonic pricing? A new multivariate approach. *International Journal of Tourism Research*, 22(3), 277–288.
3. Webb, T., Schwartz, Z., Xiang, Z., Altin, M. (2022). Hotel revenue management forecasting accuracy: The hidden impact of booking windows. *Journal of Hospitality and Tourism Insights*, 5(5), 950–965.

Finito di stampare
nel mese di Aprile 2025
da Enzo Albano s.r.l.
Via Enrico Fermi, 17 - Napoli

Euro 20,00

ISBN 979-12-80655-52-3