Modeling the Total Daily Revenue at the I Hotel

Benjamin Leidig[1], Harmony Pham[1], & Monte Thomas[1]

[1] UIUC

Author Note

The authors made the following contributions. Benjamin Leidig: Formal analysis, Visualization, Writing - original draft, Writing - review & editing; Harmony Pham: Conceptualization, Investigation, Methodology, Writing - original draft, Writing - review & editing; Monte Thomas: Formal analysis, Methodology, Writing - original draft, Writing - review & editing.

Abstract

This project aims to forecast daily hotel total revenue at the I Hotel and Illinois Conference Center using time series analysis, with an intended focus on improving operational planning and revenue management. Using historical revenue data from April 2024 to March 2025, we evaluated various statistical models to account for trends, seasonality, and volatility inherent in hotel revenue streams; based on AIC, BIC, and residual diagnostics, our analysis ultimately identified a seasonal autoregressive moving average (SARMA) model with significant seasonal and non-seasonal orders of $(1,2)\times(1,1)_7$ as the best performing model for daily revenue, while a separate analysis on variance yielded a GARCH model of GJR-GARCH(1,1)-AR(7). Previously, it was believed seasonality would present itself on a monthly basis, with months like December having the highest revenue days overall. However, the analysis demonstrated only weekly seasonality, meaning that the day of the week was a more important predictor in modeling total daily revenue. The SARMA model was used to forecast a five future values and their respective prediction intervals, effectively capturing the weekly seasonality and trends in total revenue, while the GARCH model was used to demonstrate patterns in volatility. These models provide actionable insights for hotel managers to anticipate revenue patterns, allocate resources efficiently, and fine-tune pricing strategies based on seasonal trends and predicted volatility, while also establishing a methodological foundation for integrating more complex models like SARIMAX in future work. By effectively representing seasonal trends in revenue streams and future volatility through statistical methods, the project highlights the utility of data-driven decision making in the hospitality industry and its potential increase in future use.

Modeling the Total Daily Revenue at the I Hotel

1. Introduction

Predicting hotel revenue is essential for effective operational planning and strategic decision-making. This project aims to develop a predictive model using time series analysis to forecast future revenue at the iHotel and Illinois Conference Center, a premier hotel and event venue located near the University of Illinois campus and Research Park. The facility offers 125 hotel rooms and more than 70,000 square feet of conference space, making it a popular location for university-sponsored conferences (Conference Center, 2025). In 2024, the iHotel achieved an occupancy rate of 82%, significantly higher than the national average of 65.1% and Chicago's average of 67.6% (HospitalityNet, 2024). This high occupancy rate is often driven by university-related events, including sports games, academic activities, and special celebrations. Popular events such as game days and Mom's or Dad's weekends are typically scheduled and fully booked a year in advance. However, a high occupancy rate alone does not fully reflect operational efficiency; therefore, monitoring revenue is essential to maximizing profitability. The objective of this analysis is to apply time series modeling techniques to one year of historical revenue data in order to generate accurate forecasts for future periods. The dataset comprises 365 daily observations of total hotel revenue from March 28, 2024, to March 27, 2025. Each observation includes the date, the corresponding weekday, and the total revenue in U.S. dollars. Revenue figures range from $3,364 to $52,518, with a median of $16,688 and a mean of $16,466, indicating a moderately right-skewed distribution. As illustrated in the histogram, the majority of daily revenues fall between $10,000 and $25,000, while a smaller number of days exceed $40,000. This pattern likely reflects occasional spikes in demand associated with special events or conferences. In this context, time series models such as SARIMA and GARCH are well-suited for capturing underlying trends, seasonal variations, and random fluctuations present in hotel revenue data.

2. Methods

2.1 *Determining Stationarity*

Daily total revenue was plotted over time (represented in days) and the plot was examined for stationary behavior. Visual examination included inspecting whether the variance remained relatively constant over time and if the mean function remained relatively constant over time. Inspecting the mean function also included investigating obvious trends. Alongside visual inspection, more rigorous statistical tests were used to conclude stationarity. An Augmented Dickey-Fuller (ADF) test was conducted, with the null and alternative hypotheses being:

$$H_0 : \text{The time series has a unit root.}$$

$$H_1 : \text{The time series doesn't have a unit root.}$$

Effectively, this means the null hypothesis represents non-stationarity whereas the alternative hypothesis represents stationarity. This test was conducted with a significance level of 0.05. Thus, if the test yielded a p-value below 0.05, it supports the conclusion that the series is stationary. Ultimately, however, determining stationarity wasn't definitively concluded using statistical testing alone, although this method provided useful insights and supported the conclusion made by visual inspection.

Figure 1 demonstrated a relatively stable mean function and non-constant variance. The mean function appeared to increase in the month of October, 2024, and decreased into the months of December, 2024, and January, 2025. The overlaid trend on the time series plot exemplifies the lack of a consistent, non-zero, linear trend in the time series, which suggested that differencing or detrending the data would be unnecessary. The variance function increased drastically in the months of May, September, and November, 2024.

An ADF test yielded a p-value below the 0.05 significance level, which supported the conclusion that this time series was stationary.

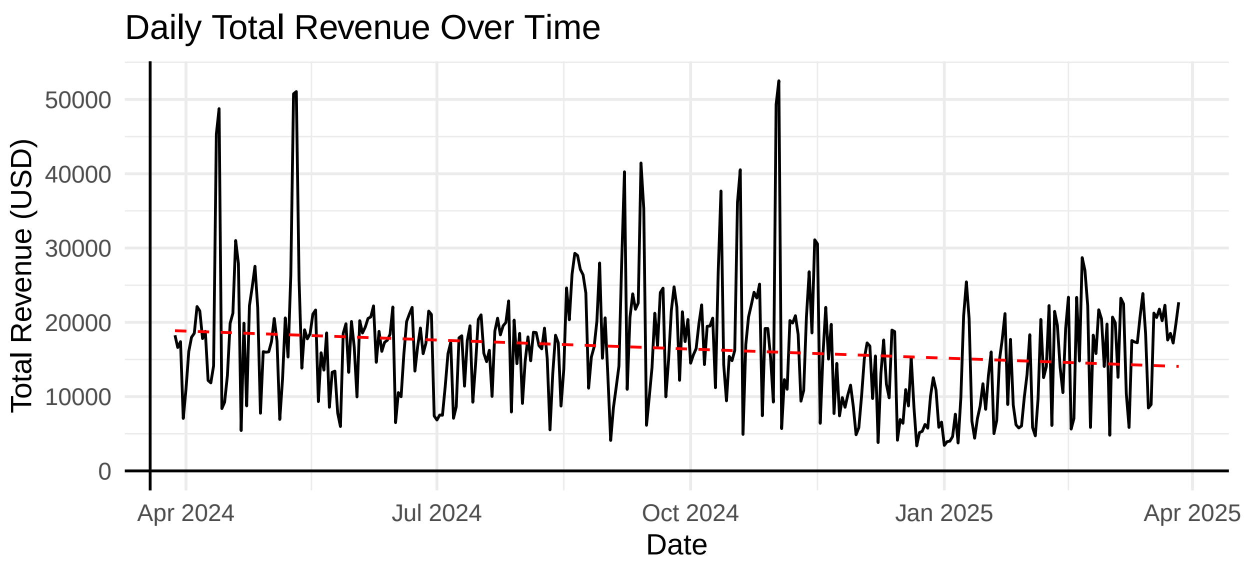

Figure 1. Daily Total Revenue Over Time

2.2 *Autocorrelation Analysis and Preliminary Model Identification*

The autocorrelation function (ACF, $\rho(h)$) was then examined for significance at various lags, h, where statistical significance at the 95% confidence interval is determined approximately by $\pm\frac{2}{\sqrt{n}}$. This inspection was carried through until $h = 100$. To determine the presence of seasonality in the time series, the ACF was also examined for significance and/or patterns at seasonal lags, $h = sk$, where $k \in \mathbb{Z}^+$, k is a seasonal lag, and s is the seasonal component. Significant ACF values at multiple lags of h in intervals of length s were used to determine seasonality and, if seasonal, to determine the seasonal component s.

The partial autocorrelation function (PACF, ϕ_{hh}) was also examined for significance at various lags, h, where statistical significance at the 95% confidence level is determined approximately by $\pm\frac{2}{\sqrt{n}}$, and performed until $h = 100$. Again, the PACF was also inspected at seasonal lags, k, as well. To determine potential orders for various candidate models, the ACF and PACF were again examined.

It can be shown that for autoregressive models of order p (AR(p)), the ACF "tails out" to insignificance while the PACF cuts off (abrupt insignificance) for lags greater than

p $(h > p)$ (see Equation 1). Similarly, for pure seasonal autoregressive models of order P (SAR(P)), the ACF tails out while the PACF cuts off for seasonal lags greater than P $(k > P)$ (Shumway & Stoffer, 2000).

The inverse is true for moving-average models of order q (MA(q))–that is, the ACF cuts off for lags greater than q while the PACF tails out (see Equation 2). Similarly, for pure seasonal moving-average models of order Q (SMA(Q)), the ACF cuts off for seasonal lags greater than Q while the PACF tails off (Shumway & Stoffer, 2000).

However, for autoregressive moving-average models of orders p and q (ARMA(p, q)) and pure seasonal autoregressive moving-average models of orders P and Q (SARMA(P, Q)), both the ACF and PACF tail out (Shumway & Stoffer, 2000). Accordingly, the behavior of the ACF and PACF of the time series was analyzed and several candidate models of different orders were fitted. Since ARMA and pure SARMA model orders cannot be determined precisely via ACF and PACF plot inspection, the orders were ambiguous, and several candidate models were created with various orders.

Figure 2 exhibited the ACF and PACF of the time series. The various ACF and PACF values for different lags can be interpreted in several ways.

One interpretation was that the ACF spiked at lag $h = 1$, and cut off to 0 for non-seasonal lags afterward. The ACF at seasonal lags tailed out to insignificance. The PACF at non-seasonal lags and seasonal lags tailed out to insignificance. Due to this behavior, the non-seasonal orders can be interpreted as $(\phi = 0, \theta = 1)$, or MA(1) model behavior. The seasonal orders can be interpreted as $(\Phi \geq 1, \Theta \geq 1)_{s=7}$.

A second interpretation was that the ACF at non-seasonal and seasonal lags tailed out to insignificance. The PACF at non-seasonal lags and seasonal lags also tailed out to insignificance. Due to this behavior, the non-seasonal orders can be interpreted as $(\phi \geq 1, \theta \geq 1)$, or ARMA$(\geq 1, \geq 1)$ model behavior. The seasonal orders can be interpreted as $(\Phi \geq 1, \Theta \geq 1)_{s=7}$.

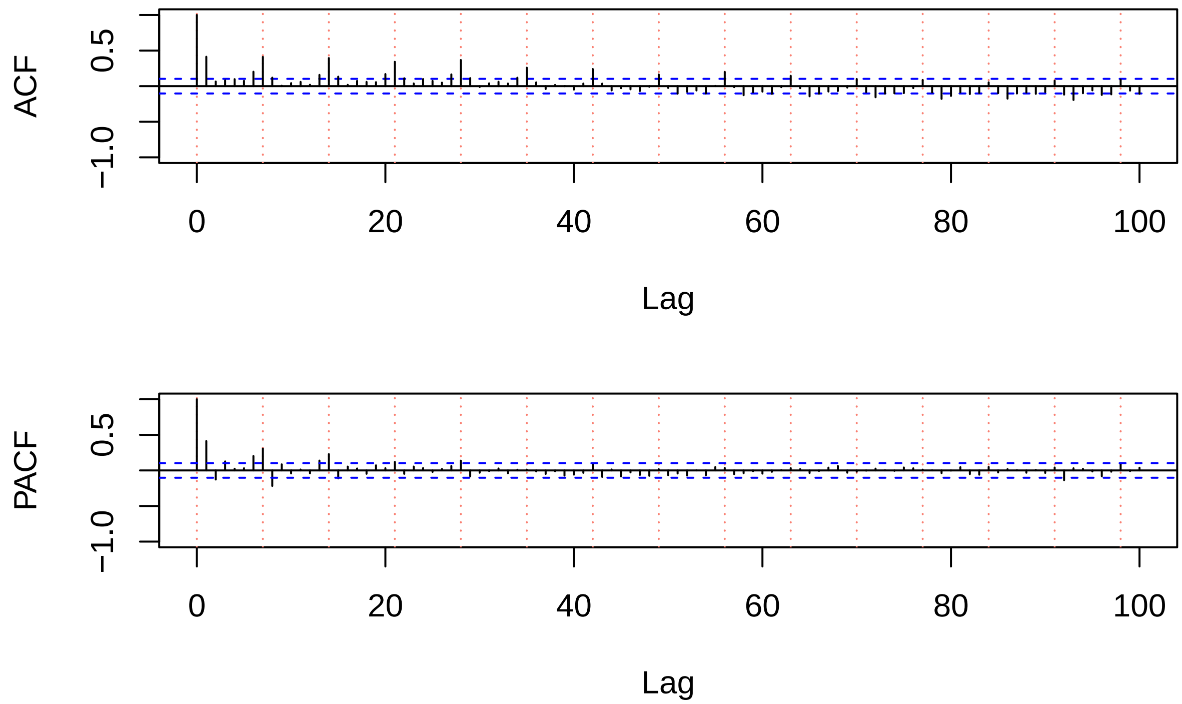

Figure 2. ACF and PACF Plots

The models of both interpretations were tested with ambiguous orders set to several combinations of 1, 2, and 3 to investigate whether higher orders were necessary in the final model. Each candidate model was investigated for parameter significance of each parameter estimate at the 95% confidence level. Models with insignificant parameter estimates were disregarded. Models with p, P, and Q orders greater than 2 all had insignificant parameter estimates for the associated order's parameter. No parameter of order 3 or greater from any model was deemed statistically significant.

2.3 ***SARMA Model Screening***

Accordingly, three models were selected for the final screening phase, each with statistically significant parameter estimates: SARMA($(0,1) \times (1,1)_7$), SARMA($(1,1) \times (1,1)_7$), and SARMA($(1,2) \times (1,1)_7$). The models kept after the first screening process were then compared using AIC, AICc, and BIC scores, the lowest scores of which are preferred. The candidates were then ranked according to these scores, and the model with the lowest BIC was chosen.

The three aforementioned models were tested using AIC, AICc, and BIC scores in Table 3. SARMA$((1,2) \times (1,1)_7)$ obtained the lowest BIC, SARMA$((0,1) \times (1,1)_7)$ obtained the second lowest BIC, and SARMA$((1,1) \times (1,1)_7)$ obtained the highest BIC. Not only did SARMA$((1,2) \times (1,1)_7)$ obtain the lowest BIC, but it also obtained the lowest AIC and AICc, too. Thus, SARMA$((1,2) \times (1,1)_7)$ was chosen as the final model.

2.4 *SARMA Residual Analysis*

To ensure robustness of analysis, a residual analysis was conducted on the chosen model, including a Ljung-Box test with the respective null and alternative hypotheses:

$$H_0 : \text{The residuals don't exhibit autocorrelation.}$$

$$H_1 : \text{The residuals exhibit autocorrelation.}$$

So, p-values above 0.05 are preferred, as an assumption in the modeling is uncorrelated, normally distributed errors with mean zero ($W_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2)$).

Figure 7 showed all p-values from the Ljung-Box test to be greater than a significance level of 0.05, failing to reject the null hypothesis that the residuals don't exhibit autocorrelation. So, each observed residual from the model is uncorrelated with every other observed residual. Figure 8 supported this conclusion, as the ACF of residuals displayed insignificant autocorrelation at any lag $h > 0$. It's important to note that this demonstrated not only uncorrelated errors, but a lack of seasonality in residuals as well. This means seasonal dependencies were correctly incorporated into the final model. Alongside this test, the residuals were plotted over time to ensure a zero mean and constant variance. Similarly to how the original time series was investigated, an ADF test was conducted to support the conclusion made on the nature of the residuals in order to conclude zero mean and constant variance, abiding by the aforementioned stipulation of the model assumption. The standardized residuals of the chosen model appeared to have a constant mean function in Figure 9. However, the variance of the residuals fluctuates mildly, increasing near the start (April, 2024) and middle (October, 2024) of the series. However, an ADF test yielded a p-value below the 0.05 significance level, which suggested the residuals were stationary and thus had constant mean and variance. The theoretical normal and sample quantiles aligned for most observations in the dataset, except for a few irregularities for quantiles greater than or less than 4 in Figure 10. Therefore, the residuals of the model were normally distributed with mean zero, which follows the model assumption $W_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2)$.

2.5 *Modeling Heteroskedasticity*

The application of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models in the hospitality and tourism industry has gained prominence as a tool for understanding volatility and forecasting demand. Tang et al. (Tang, Ramos, Cang, and Sriboonchitta (2017)) used a copula-GARCH approach to analyze inbound tourism demand in China. Their study revealed that tourist arrivals exhibit significant seasonal patterns and tail dependence, meaning extreme events lead to synchronized fluctuations in arrivals from different countries. The incorporation of copulas allowed the researchers to capture nonlinear dependencies that traditional linear correlation measures overlook. This offered more nuanced insights for tourism managers aiming to mitigate risks from demand shocks. Srivastava et al. (Srivastava, Chandrakala, and Suresh (2022)) extended GARCH modeling to assess the effects of the COVID-19 pandemic on hospitality stock returns. Using weekly returns from India's hospitality sector during the 2020 lockdown, they found significant volatility increases driven by pandemic-related disruptions. Their analysis demonstrated that GARCH models effectively captured time-varying volatility during crisis periods, providing a framework for investors and policymakers to understand how external shocks influence financial performance in the hospitality industry. Importantly, their findings highlight the vulnerability of hospitality stocks to public health crises and the need for volatility forecasting in strategic planning. Ampountolas (Ampountolas (2021)) compared the forecasting accuracy of several time series models, including SARIMAX, artificial neural networks, and GARCH variants, for daily hotel demand in a U.S. metropolitan market. While SARIMAX models with exogenous variables (e.g., temperature, holidays, competitor pricing) achieved strong predictive performance, the study found that both standard GARCH and GJR-GARCH models provided superior forecasting accuracy across multiple horizons. This suggests that volatility modeling is not only relevant in financial contexts but also plays a critical role in operational forecasting for hotel revenue management. The study reinforces the value of incorporating exogenous variables and volatility dynamics to enhance forecasting precision. Together, these studies establish GARCH models as versatile tools for both financial and operational forecasting in the hospitality sector. Whether applied to stock returns or demand volumes, GARCH and its extensions facilitate a deeper understanding of volatility patterns and enable stakeholders to plan for uncertainty. Moreover, recent research underscores the benefit of integrating external factors and nonlinear dependencies, moving beyond univariate models to approaches that reflect the complex, interconnected nature of hospitality markets.

2.6 *GARCH Modeling*

The time series analysis of hotel revenue focused on modeling both the conditional mean and conditional variance to capture revenue dynamics and volatility patterns. Initial exploratory analysis of the revenue data showed high variability, with clear spikes corresponding to high-demand days. A log transformation was applied to stabilize variance, and the ADF test confirmed stationarity of both the original and log-transformed series ($p < 0.01$), indicating no need for differencing. Examination of squared log returns revealed clusters of high volatility, providing visual evidence supporting the use of GARCH-type models. The initial GARCH(1,1) model, estimated using log returns, provided a basic fit to the data but failed several diagnostic checks. The sign bias test indicated significant positive sign bias ($t = 2.336, p = 0.020$) and a significant joint effect ($\chi^2 = 10.638, p = 0.013$), suggesting that the model underestimated volatility following positive shocks. Furthermore, goodness-of-fit tests across multiple groupings rejected the null hypothesis of correct specification ($p < 0.01$), and standardized residuals showed remaining autocorrelation. These findings indicated misspecification in the variance process and prompted consideration of asymmetric volatility models. A GJR-GARCH(1,1) model was subsequently estimated to account for potential asymmetry in the volatility response to shocks. The inclusion of an AR(7) term in the mean equation was motivated by weekly patterns identified in the autocorrelation function. The GJR-GARCH(1,1)-AR(7) model using normal errors reduced the sign bias (positive sign bias $t = 2.335, p = 0.021$), but the joint effect remained significant ($\chi^2 = 12.498, p = 0.006$). To account for heavy-tailed behavior in the return distribution, the model was re-estimated using a Student-t distribution for the innovations. The final model, GJR-GARCH(1,1)-AR(7) with Student-t errors, achieved the lowest AIC (1.2002) and the highest log-likelihood (-205.4) across all models tested, indicating superior fit.

2.7 *GARCH Residual Analysis*

Diagnostic evaluation of the final model confirmed substantial improvement in specification. The sign bias test was no longer significant for any component (Sign Bias: $t = 0.542, p = 0.588$; Negative Sign Bias: $t = 0.124, p = 0.902$; Positive Sign Bias: $t = 0.067, p = 0.947$; Joint Effect: $\chi^2 = 1.094, p = 0.778$), suggesting that the model adequately captured asymmetric volatility effects. Goodness-of-fit tests further supported model adequacy, with none of the test groupings rejecting the null hypothesis at $\alpha = 0.05$ ($p = 0.071, 0.192, 0.271, 0.251$ for group sizes 20, 30, 40, 50, respectively). Residual diagnostics showed no significant autocorrelation in standardized or squared standardized residuals, and the empirical density of residuals closely aligned with the fitted Student-t distribution, validating the model's ability to capture heavy tails.

3. Results

3.1 *SARMA Parameter Estimation*

Table 1

Model Estimates

Parameter	Estimate	SE	t-value	p-value
$\hat{\phi}$	0.9699	0.0198	48.9615	0
$\hat{\theta}_1$	-0.4487	0.0519	-8.6527	0
$\hat{\theta}_2$	-0.4131	0.0480	-8.6025	0
$\hat{\Phi}$	0.9983	0.0034	289.4971	0
$\hat{\Theta}$	-0.9583	0.0401	-23.8717	0

The parameters for the final model were estimated using the exact likelihood maximization method. All of the parameter estimates for the chosen model were significant, as shown in Table 3. Thus, the time series of total daily revenue, $\{X_t\}$, can be most accurately modeled by:

$$X_t = 0.9699X_{t-1} + 0.9983X_{t-7} + W_t - 0.4487W_{t-1} - 0.4131W_{t-2} - 0.9583W_{t-7}.$$

3.2 *SARMA Forecasting*

To forecast total daily revenue m days ahead, the *m-step-ahead* forecasting method was used. If a forecast of m-steps-ahead given a history of n previous observations is notated $\tilde{X}^n_{n+m} = E[X_{n+m}|X_n, X_{n-1}, ...]$, then Equation **??** follows. Before a forecast is made, the given model equation must be solved for its invertible representation, or it MA(∞) representation, coefficients, π_j, given by $W_t = \pi(B)X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$, where $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$, and $\sum_{j=0}^{\infty} |\pi_j| < \infty$ with $\pi_0 = 1$. Using the invertible representation and assuming infinite history, the *m-step-ahead* forecast can be found by

$\tilde{X}^n_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{X}^n_{n+m-j} - \sum_{j=m}^{n+m-1} \pi_j X_{n+m-j} - \sum_{j=n+m}^{\infty} \pi_j X_{n+m-j}$. Truncating this forecast to use only the m records observed, $\tilde{X}^n_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{X}^n_{n+m-j} - \sum_{j=m}^{n+m-1} \pi_j X_{n+m-j}$. Practically speaking, the *m-step-ahead* forecast can be obtained iteratively using the 1-step-ahead forecast until the $(m-1)$-step-ahead forecast, that is, assuming $W_0 = 0$ and $W_{n+j} = 0$ for $j = 1, ..., m$.

Five future values were forecasted according to the chosen model using the aforementioned truncated *m-step-ahead* forecasting method. The full time series plus the five forecasted future values were plotted.

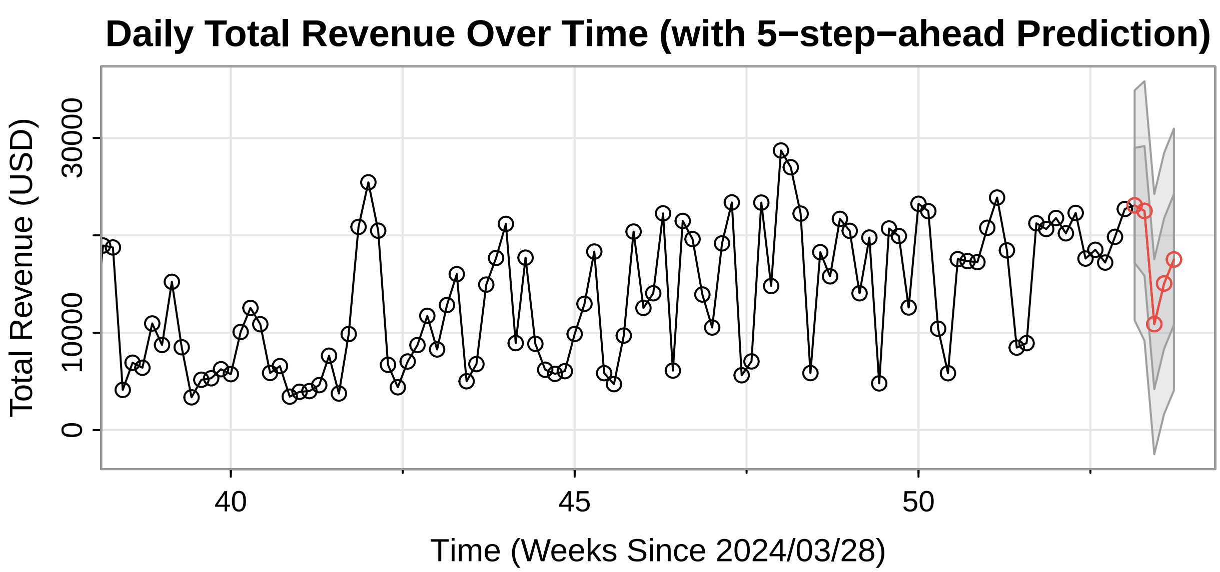

Figure 3. 5-step-ahead Forecast for Daily Total Revenue

Figure 3 showed the future five forecasted values for $\{X_t\}$, total daily revenue, using the truncated *m-step-ahead* forecasting method. The dark gray interval represents the 95% prediction interval, and the light gray interval represents the 80% prediction interval. The predicted values are shown in Table 2 as well as the respective standard errors for each prediction.

Table 2

Forecasted Values

Date	Forecast	SE
2025/03/28	23079.84	5905.305
2025/03/29	22511.09	6659.038
2025/03/30	10885.84	6681.279
2025/03/31	15057.93	6702.128
2025/04/01	17524.90	6721.676

GARCH Forecasting

Forecasting results from the final model projected relatively stable revenue over the next 10 days, with the mean forecast remaining near recent observed levels. The 95% confidence intervals gradually widened over the forecast horizon, reflecting increasing uncertainty typical of volatility models, but no sharp deviations were predicted. The forecast plot visually confirmed that forecasted volatility returned toward a steady-state level consistent with mean-reverting properties of the GARCH process (see Figure 4). Overall, the analysis demonstrates that hotel revenue exhibits volatility clustering and asymmetric responses to positive and negative shocks, with positive revenue shocks contributing more substantially to volatility increases. The final GJR-GARCH(1,1)-AR(7) model with Student-t errors successfully accounted for these dynamics, providing well-specified volatility forecasts and a reliable representation of the underlying time series process. Figure 12 shows forecasts and confidence intervals according to the GJR-GARCH(1,1)-AR(7) model predictions. Table 4 shows the exact values and bounds for each forecast. Future research may explore incorporating exogenous predictors such as holidays, promotional campaigns, or local events to improve forecasting accuracy and provide actionable insights for revenue management.

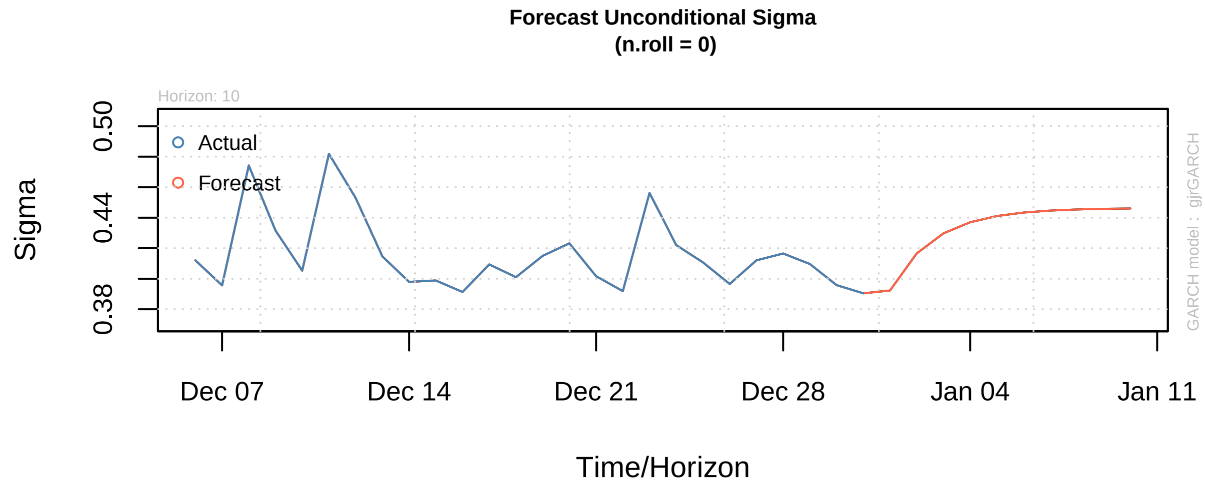

Figure 4. Forecasted Variance

Discussion

In this project, two time series models were identified and analyzed to forecast daily total revenue for the iHotel and Illinois Conference Center. The main goal was to develop a reliable forecasting model that could support hotel management decisions, including pricing, budgeting, and resource planning. Based on the earlier analysis, two models were selected: SARMA$((1,2) \times (1,1)_7)$ and GJR-GARCH(1,1). This discussion compares their performance and explains the strengths and weaknesses of each model in capturing trend, seasonality, and volatility. The SARMA$((1,2) \times (1,1)_7)$ model, a simplified version of SARIMA without differencing, effectively captured both non-seasonal and seasonal patterns with a 7-day period, as identified through ACF and PACF analysis. This model was chosen after comparing AIC, AICc, and BIC scores among several candidates. All parameter estimates were statistically significant, and the residuals showed no autocorrelation based on the Ljung-Box test. The model successfully forecasted five future revenue values, ranging from \$10,885.84 to \$23,079.84, with both 80% and 95% prediction intervals. Although the SARMA model passed the ADF test and showed constant mean and variance overall, the residuals still had some small changes in variance, particularly in May, September, and November 2024. This suggests that the model might not fully capture changes in volatility during certain times, which could make the forecasts less reliable in those periods.

To better model the variance in revenue returns, the GJR-GARCH(1,1) model was used. Initial analysis of log returns and their squared values showed volatility clustering, supporting the use of GARCH-type models. The standard GARCH(1,1) model provided a basic fit but failed diagnostic tests, showing significant positive sign bias and joint effects. To address these issues, a GJR-GARCH(1,1) model was estimated with an AR(7) mean component and Student-t innovations to account for both weekly seasonality and heavy tails. This final model achieved the lowest AIC and highest log-likelihood among all models tested and passed all diagnostic checks, including sign bias and residual autocorrelation tests. Forecasts from this model projected daily revenue ranging from $18,996.72 to $21,331.87, with 95% confidence intervals spanning from as low as $8,028.88 to as high as $50,996.48, effectively capturing the potential uncertainty and volatility observed in actual revenue trends.

A comparison between forecasted and actual revenue from March 28 to April 5 revealed notable differences in model performance. While the SARMA$((1,2) \times (1,1)_7)$ model performed well during relatively stable periods, accurately capturing weekly trends and short-term fluctuations, it consistently underestimated revenue spikes on high-demand days such as April 4 and April 5, when actual revenue reached more than $48,798 and $49,795, respectively. In contrast, the GJR-GARCH(1,1) model, although slightly underpredicting those peak values, provided broader prediction intervals that successfully encompassed the actual outcomes. This reflects the GJR-GARCH model's advantage in capturing heteroskedasticity and responding adaptively to sudden changes in revenue patterns.

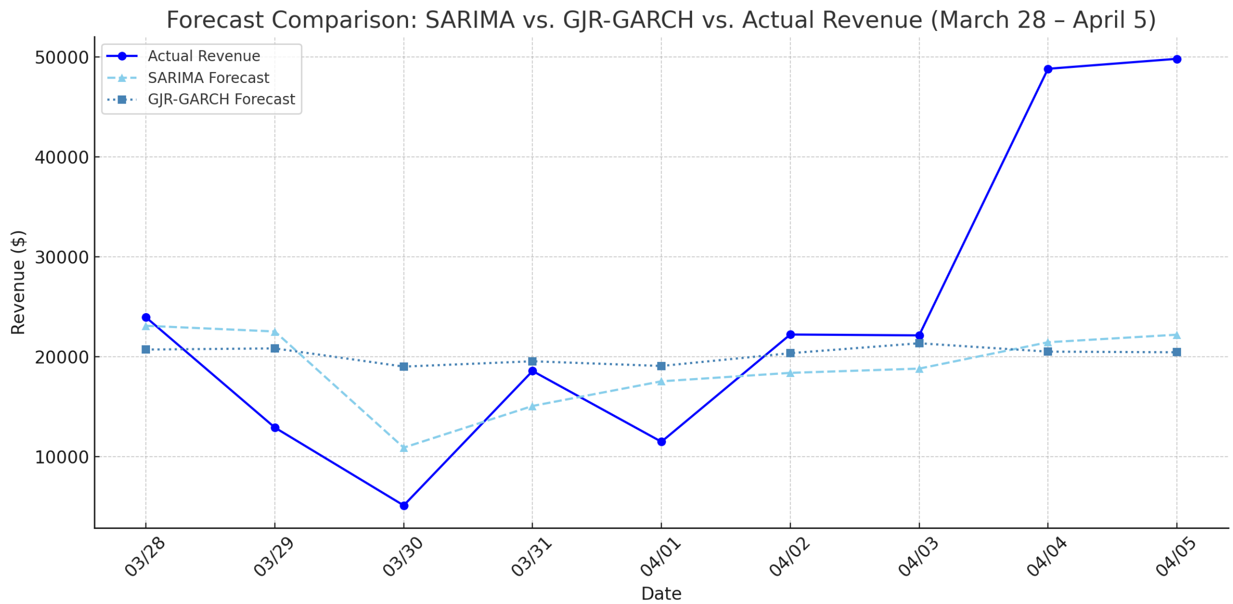

Figure 5. SARMA & GJR-GARCH Forecasts versus Actual Revenue

Overall, both models have their strengths. The SARMA model is strong at capturing the structure of the iHotel's revenue data, including trend and seasonality, and it performs well for point forecasting. However, it is less effective in modeling changes in variance. On the other hand, the GJR-GARCH model focuses on the volatility of the data and captures time-varying variance and asymmetric effects more effectively, which are important when revenue experiences large or sudden changes. A combination of both approaches or a hybrid model may offer even better forecasting performance in future research.

The results of this study directly support its main goal that is to build a reliable model for forecasting daily hotel revenue using historical data. The selected SARMA$((1,2) \times (1,1)_7)$ model successfully captured both short-term dependencies and weekly seasonality, patterns that are common in hotel operations. This model generated five-day forecasts, providing hotel managers with both point estimates and a range of potential outcomes. This helps improve short-term financial planning and allows the iHotel to better prepare staffing and pricing strategies based on expected demand.

The findings also strongly reflect the motivation for this analysis. The iHotel, located near a university campus and research park, often experiences revenue fluctuations tied to local events such as academic conferences, company events, sports games, graduations, and family weekends. Although these outside factors were not directly included in the model, the weekly seasonal pattern we found probably reflects their impact. This study gives a strong starting point for building forecasting tools that can help the iHotel make better decisions, especially when demand changes due to local events. In the future, the model can be improved by adding event dates and other related information to make the forecasts more accurate.

Despite the strong results, this study still has some limitations. First, it used only total daily revenue as the time series variable. Other useful factors such as ADR (average daily rate), occupancy rate, special events, or holidays were not included. Second, the dataset covered only one year. This short time period limits the model's ability to detect long-term patterns or rare seasonal changes, such as holiday surges or large university events. Finally, although the model performed well on the available data, it was not tested on data from other years. Without cross-validation, we cannot be sure how well the model will work in the future. For future work, predictive accuracy could be further enhanced by incorporating models such as SARIMAX and Facebook Prophet, particularly when hotel revenue is influenced by external factors. SARIMAX extends the traditional SARIMA model by including exogenous variables, such as holidays, local events, and marketing campaigns, making it especially suitable for settings where occupancy is driven by external influences (Saputra, Adikara, & Wirawan, 2023). Meanwhile, Prophet, developed by Meta, is an open-source forecasting tool designed to capture seasonality, holidays, and trend shifts with minimal manual tuning. It is particularly effective for daily or irregular time series data commonly found in hotel bookings (Taylor & Letham, 2017).

References

Ampountolas, A. (2021). Modeling and forecasting daily hotel demand: A comparison based on SARIMAX, neural networks, and GARCH models. *Forecasting*, *3*(3), 580–595. https://doi.org/10.3390/forecast3030037

Conference Center, iHotel &. (2025). iHotel and illinois conference center. Retrieved from https://stayatthei.com/

HospitalityNet. (2024). STR: U.s. Hotel results for week ending 2 march. Retrieved from https://www.hospitalitynet.org/news/4127014.html

Saputra, M. W. N., Adikara, I. N. G., & Wirawan, D. G. A. (2023). Forecasting hotel occupancy rates in bali province using the SARIMAX method with tourist data as an exogenous variable. Retrieved from https://www.researchgate.net/publication/385675519

Shumway, R. H., & Stoffer, D. S. (2000). *Time series analysis and its applications.* Springer.

Srivastava, P., Chandrakala, D. P., & Suresh, N. (2022). GARCH model for determining COVID-19 pandemic effect on hospitality stock returns. *Proceedings of the 2nd Indian International Conference on Industrial Engineering and Operations Management*, 970–973.

Tang, J., Ramos, V., Cang, S., & Sriboonchitta, S. (2017). An empirical study of inbound tourism demand in china: A copula-GARCH approach. *Journal of Travel & Tourism Marketing*, *34*(9), 1235–1246. https://doi.org/10.1080/10548408.2017.1330726

Taylor, S. J., & Letham, B. (2017). Forecasting at scale. *PeerJ Preprints.* https://doi.org/10.7287/peerj.preprints.3190v2

Appendix

ACF and PACF of AR(p):

$$
\begin{aligned}
\rho(h) &= \begin{cases} 1, & h = 0 \\ \sum_{i=1}^{p} \phi_i \rho(h-i), & h \geq 1 \end{cases} \\
\hat{X}_{t+h} &= \sum_{j=1}^{p} \phi_j X_{t+h-j} \\
\phi_{hh} &= \mathrm{Corr}(X_{t+h} - \hat{X}_{t+h},\ X_t - \hat{X}_t) \\
&= \mathrm{Corr}(W_{t+h},\ X_t - \hat{X}_t) = 0, \quad h > p
\end{aligned} \tag{1}
$$

ACF and PACF of MA(q):

$$
\rho(h) = \begin{cases} \frac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{1+\sum_{i=1}^{q} \theta_i^2}, & 0 \leq h \leq q \\ 0, & h > q \end{cases} \tag{2}
$$

$$
\phi_{hh} = \begin{cases} 1, & h = 0 \\ -\frac{(-\theta)^h (1-\theta^2)}{1-\theta^{2(h+1)}}, & h \geq 1 \end{cases}
$$

$$
\tilde{X}^n_{n+m-j} = \begin{cases} \tilde{X}^n_{n+m-j}, & j < m \\ X^n_{n+m-j}, & j \geq m \end{cases} \tag{3}
$$

Table 3

Model Performance

Orders	**AIC**	**AICc**	**BIC**
(1,2)×(1,1)$_7$	20.28101	20.28165	20.35580
(0,1)×(1,1)$_7$	20.31236	20.31266	20.36578
(1,1)×(1,1)$_7$	20.30773	20.30818	20.37183

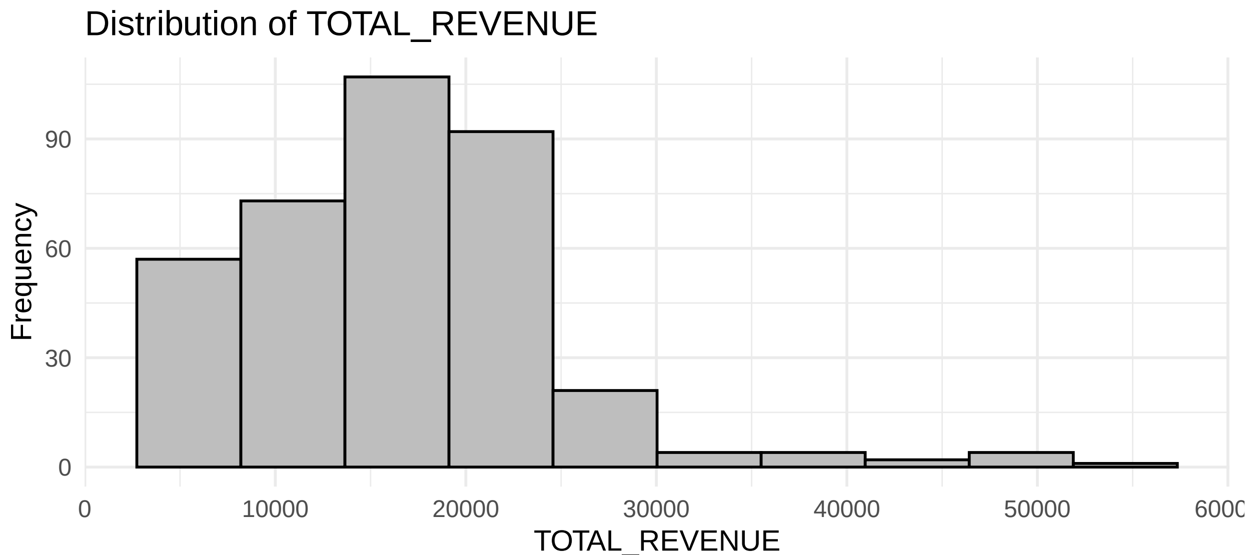

Figure 6. Histogram of Total Daily Revenue

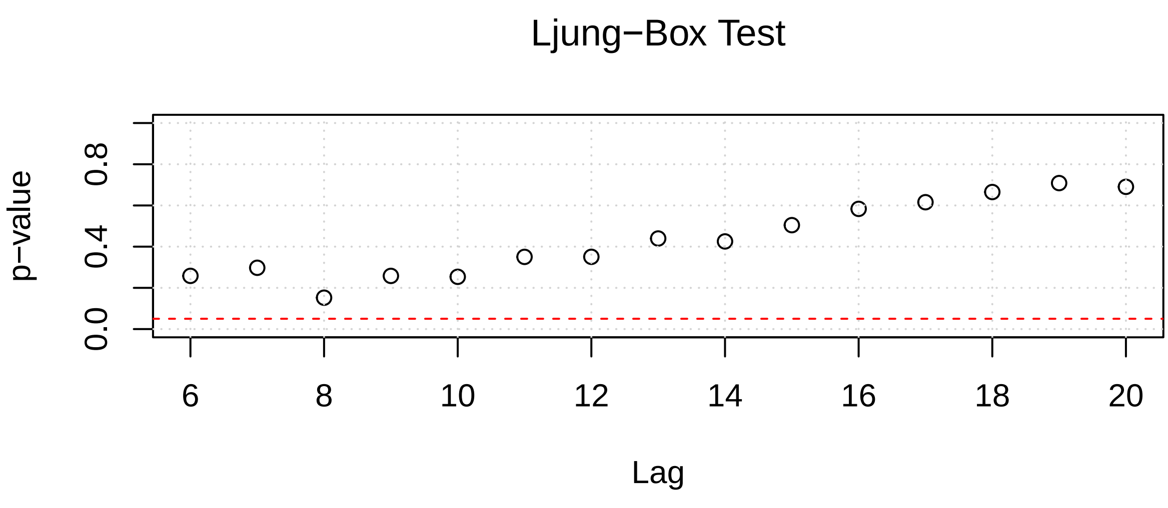

Figure 7. Ljung-Box Test on the SARMA Model Residuals

Table 4

GARCH Model Forecasts

T+	Date	Forecasted Revenue	Lower	Upper
1	2025-03-28	20702.85	9596.63	44662.35
2	2025-03-29	20817.27	9200.71	47100.60
3	2025-03-30	18996.72	8182.28	44104.49
4	2025-03-31	19528.93	8292.37	45991.59
5	2025-04-01	19059.84	8028.88	45246.32
6	2025-04-02	20339.07	8529.41	48500.17
7	2025-04-03	21331.87	8923.14	50996.48
8	2025-04-04	20493.90	8560.36	49063.35
9	2025-04-05	20428.54	8526.16	48946.46

Table 5

Forecasted Revenue with 95% Confidence Interval from 2025-03-28 to 2025-04-05

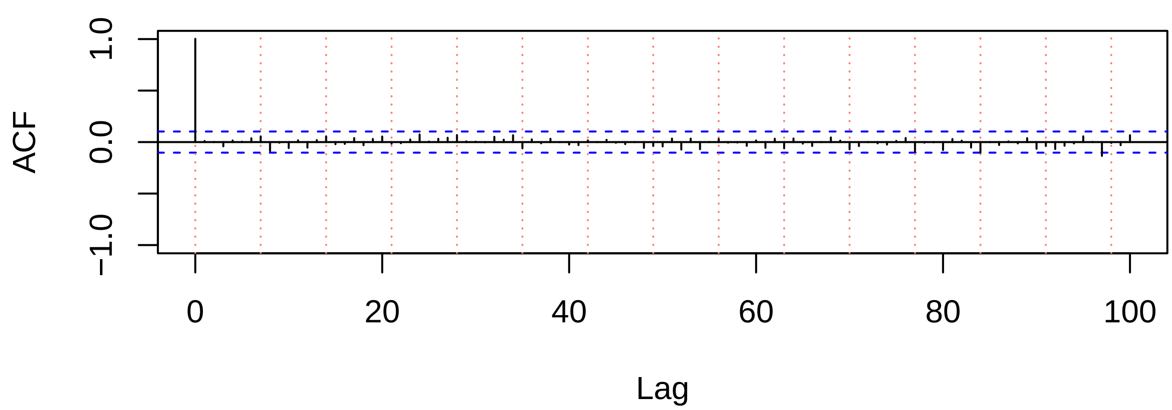

Figure 8. ACF Plot of the SARMA Model Residuals

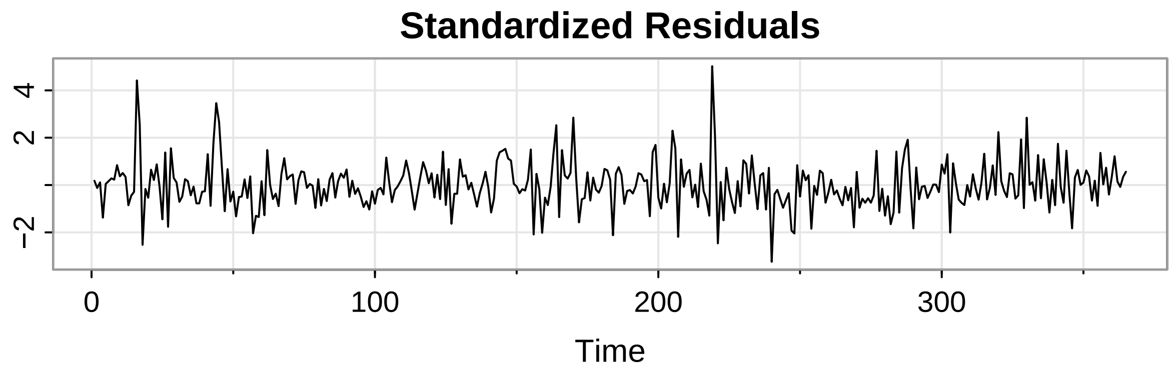

Figure 9. Standardized Residuals Over Time

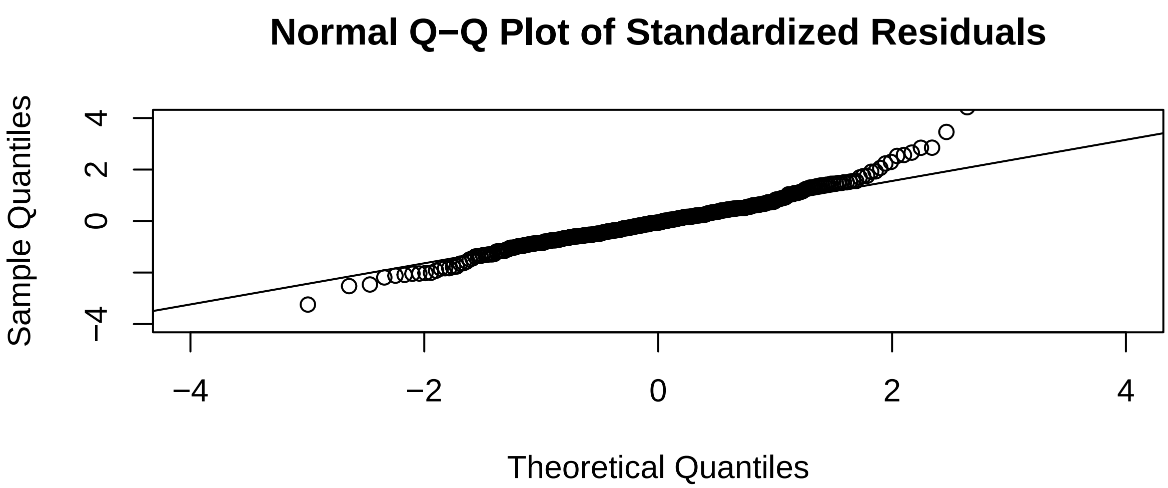

Figure 10. Normall Q-Q Plot of Standardized Residuals of the SARMA Model

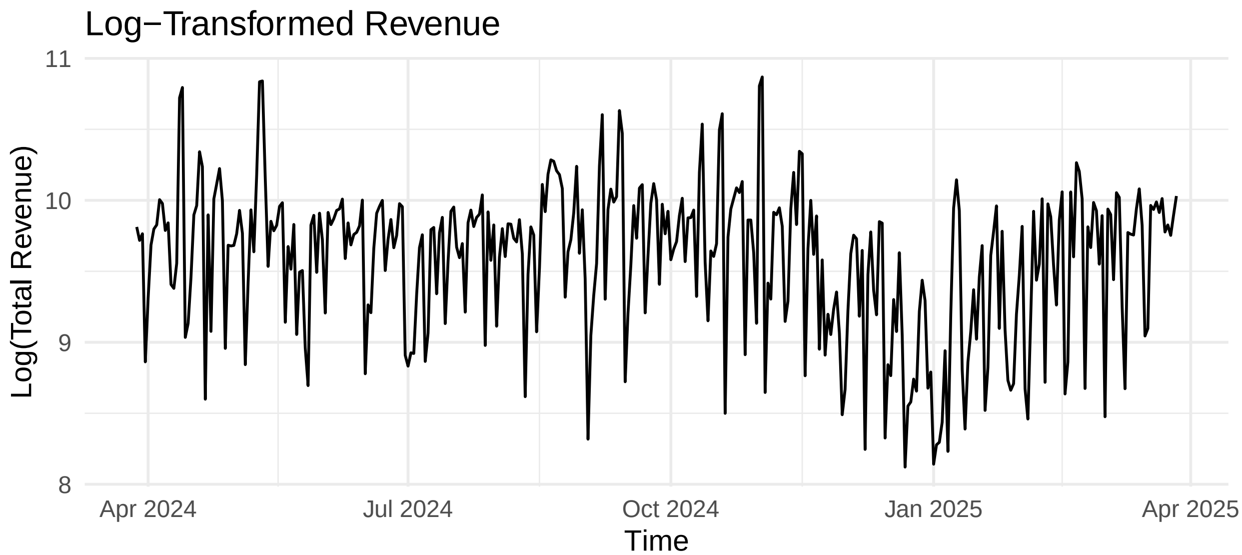

Figure 11. Log Daily Total Revenue Over Time

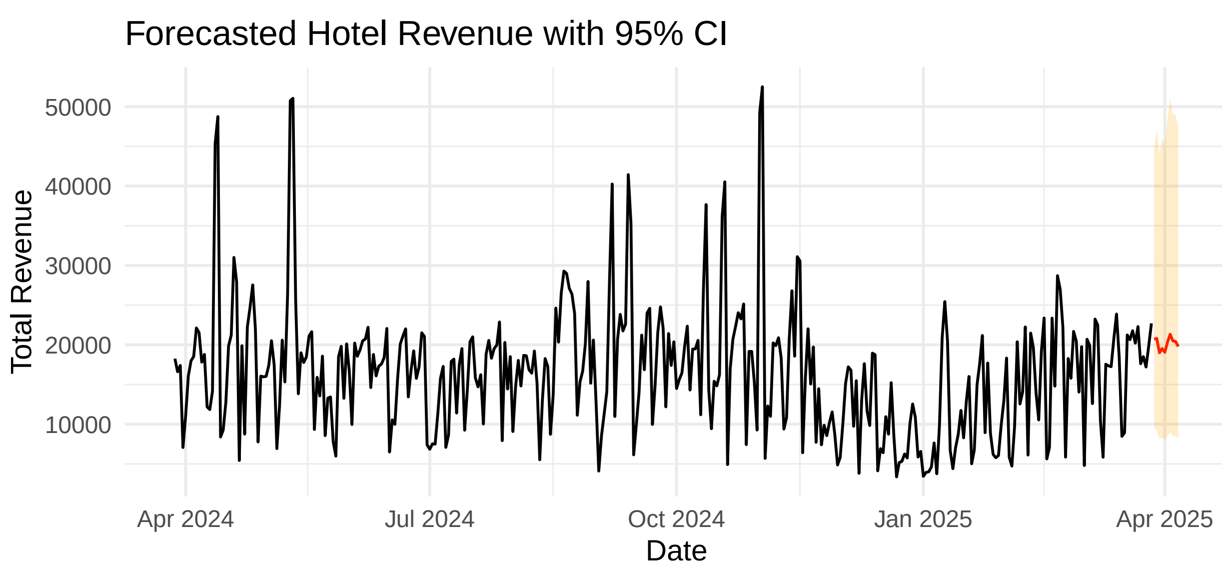

Figure 12. Forecasted Hotel Revenue Using GARCH