Forecasting Tourism in Switzerland

Antoine Magnin, Robin Michel, Pierre Monari, Ronald Medvedev, and Faber Bickerstaffe

University of Lausanne, HEC

Forecasting 2

2024

Group 3

20.05.2024

Introduction

The Swiss tourism industry plays a crucial role in the national economy, with regions like Vaud and Luzern attracting significant numbers of visitors due to their unique attractions and cultural significance. Accurate forecasting of overnight stays is helpful for effective economic planning, resource management, and strategic marketing. This report undertakes a dual approach to forecasting, addressing both the general demand for accommodation in Vaud and the specific trends concerning visitors from Japan in Luzern.

Part a) of the project focuses on predicting the total number of overnight stays in the canton of Vaud. This analysis aims to provide a broad understanding of tourism dynamics and assist local stakeholders in planning and decision-making processes. Part b) delves into a more detailed prediction for Luzern, focusing on overnight stays by Japanese tourists, a demographic that presents unique challenges due to specific preferences and travel behaviors.

Through the application of various forecasting models this project not only aims to generate reliable forecasts but also to equip tourism operators and policymakers with the tools needed to enhance service delivery and optimize marketing strategies. The inclusion of confidence intervals in our predictions will further aid stakeholders in assessing risk and making informed decisions in a sector known for its volatility.

Data

Description

The dataset provided comprises of 473850 observations and 5 columns, detailing overnight stays across various cantons spanning from 2005 to 2023. The dataset includes information from 27 different cantons and 77 countries of origin. The columns, which include Herkunftsland, Kanton, Monat, Jahr, value, provide a comprehensive view of tourist overnight stays in Switzerland. Here are the roles and significance of each column:

- `Herkunftsland`: Identifies the visitor's country of origin, critical for understanding international tourism trends.
- `Kanton`: Indicates the canton where the stays were recorded, essential for regional analysis.
- `Monat and Jahr`: Time indicators that help in time series forecasting.
- `Value`: Represents the counted overnight stays, serving as the primary metric for our forecasts.

In the appendix a random sample snapshot of the data can be found to provide a tangible sense of the data structure (see Appendix 1).

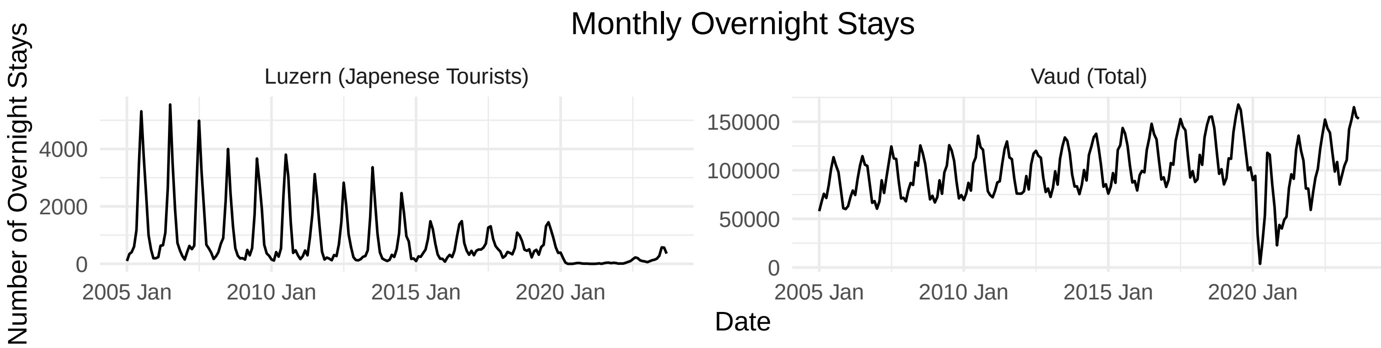
Processing Data

For part a) of the project, we focused on the Canton of Vaud, filtering the data for all overnight stays in this region. The inclusion of "Herkunftsland - Total" in the `Herkunftsland` column helped with the filtering as it provided a monthly total of visitors of all nationalities. We therefore filtered on the columns of `Kanton` and `Herkunftsland`. The data was then converted into a time series format, with the `Date` column created by combining the `Monat` and `Jahr` columns. This transformation enabled us to visualize trends and seasonality in the data, providing a foundation for our forecasting models.

For part b) of the project, we focused on Luzern and Japanese visitors, filtering the data to isolate overnight stays by Japanese tourists in this region. "Japan" is a country value in the `Herkunftsland` column and therefore could be selected for filtering. Similar to part a), the data was converted into a time series format, with the `Date` column once again created by combining the `Monat` and `Jahr` columns. This transformation allowed us to explore specific trends and patterns related to Japanese visitors in Luzern. To aid in the creation of the tsibbles for both parts of the project, we created a conversion table for the months in German to their numerical equivalent. This conversion was necessary for the correct ordering of the months in the time series data.

Analysis

The analysis of the data involved visualizing the trends and seasonality in overnight stays for both Vaud and for Japanese tourists in Luzern. This visualization was essential for understanding the patterns in the data and identifying any anomalies or trends that could inform our forecasting models. The time series plots below provide a clear representation of the overnight stays in Vaud and Luzern, highlighting the seasonal variations and trends over time.



Seasonality

- `Luzern`: From 2005 to 2019, overnight stays by Japanese tourists exhibited a consistent pattern. Seasonal peaks in overnight stays by Japanese tourists are observed in the summer months (June, July, August) and troughs in the winter months (January, February).
- `Vaud`: Similar seasonal peaks are observed for all tourists, indicating the popularity of the summer season and reduced interest in the winter, likely due to weather conditions and fewer attractions.

Trend

- `Luzern`: A downward trend has been evident since 2007, likely due to economic challenges in Japan. The highest values of overnight stays by Japanese tourists were observed from 2005 to 2007 (17,000-19,000 overnight stays per year), followed by the decline.
- `Vaud`: From 2005 to 2019, the number of overnight stays by all tourists gradually increases, especially after 2014, reaching peaks in 2018 and 2019 (around 3,000,000 overnight stays). This may indicate growing popularity of the region and improved tourist infrastructure. (see Appendix 2)

COVID-19 Pandemic

- In 2020, the number of overnight stays dropped sharply to minimal levels in both regions. In Lucerne, annual overnight stays by Japanese tourists fell to less than 7,000, and in Vaud, the number dropped to around 1,500,000 for all tourists. This is attributed to global travel restrictions and a general decline in tourist activity.
- In 2021, numbers remained low, but a gradual recovery began towards the end of the year. This could be due to the easing of restrictions and renewed interest in travel.
- By 2023, recovery continued, although peak numbers have not yet reached pre-pandemic levels, likely due to ongoing safety concerns and changing tourist preferences.

Forecating Methodology

In forecasting the trends of tourist visits in the cantons of Vaud and Luzern, we compared the suitability of various statistical models to account for the intricate dynamics and seasonality inherent in tourism data. The selection of Exponential Smoothing State Space Model (ETS), Autoregressive Integrated Moving Average (ARIMA), ARIMA with dynamic regression, and Time Series Linear Model (TSLM) is predicated on their efficacy in handling data with distinct patterns and external influences:

`ETS Models` are particularly adept at capturing seasonal variations and trends without assuming stationarity in the time series data. This makes ETS models highly suitable for our tourism data, where seasonal peaks and evolving trends are prominent due to varying tourist preferences and external economic conditions.

`ARIMA Models` are chosen for their flexibility in modeling a wide range of data behaviors through differencing, making them capable of stabilizing the non-stationary elements of our series. This model is robust in dealing with the autocorrelation within the data, allowing for precise short-term forecasts which can be crucial for operational planning in tourism management.

`ARIMA with Dynamic Regression` extends the basic ARIMA framework by incorporating exogenous variables, such as currency exchange rates and dummy variables for shock global events such as COVID19, which are presumed to influence tourist inflows. This approach is particularly useful for understanding and forecasting the impact of external economic factors on tourism, such as the YEN vs CHF exchange rate's effect on Japanese visitors in Luzern.

`TSLM` offers a regression-based approach tailored to time series data, allowing the inclusion of both seasonal dummies and trends within the model. It provides a structured way to incorporate multiple seasonal cycles and linear trends, making it invaluable for long-term strategic planning where understanding the impact of specific time events (like COVID19) is essential.

These models were chosen to address the specific challenges presented by the dataset, including the need to forecast accurately in the face of seasonal fluctuations, economic impacts, and unexpected global events. The following sections will detail the model configurations, and the rationale behind the choice of each modeling technique in the context of our objectives, and finally the selected model based off our chosen metric.

ETS Model

An automatic approach was used to identify the optimal ETS model for forecasting overnight stays in the cantons of Vaud and Luzern. The model selection process involved comparing the AIC values of different ETS models with varying error structures (additive (A), multiplicative (M)), trend structures (A, M, None (N)) and season components (A, M, N). The final ETS model was chosen based on the lowest AIC value, indicating the best fit to the data.

`VAUD Best ETS Model: ETS(A, N, A)`

`LUZERN Best ETS Model: ETS(M, N, M)`

ARIMA & ARIMA with Dynamic Regression

A stepwise approach was used to identify the optimal ARIMA model for forecasting overnight stays in the cantons of Vaud and Luzern. The model selection process involved identifying the optimal p, d, q, and P, D, Q parameters for the ARIMA model based on the AIC values. The final ARIMA model was chosen based on the lowest AIC value, indicating the best fit to the data.

`VAUD Best ARIMA Model: ARIMA(1,0,1)(2,1,2)[12]`

Non-seasonal part: 1 autoregressive term, no differencing, and 1 moving average term. Seasonal part: 2 seasonal autoregressive terms, 1 order of seasonal differencing, and 2 seasonal moving average terms, all based on a yearly cycle (12 months).

`LUZERN Best ARIMA Model: ARIMA(1,0,0)(2,1,1)[12] w/ drift`

Non-seasonal part: 1 autoregressive term, no differencing, and no moving average terms. Seasonal part: 2 seasonal autoregressive terms, 1 order of seasonal differencing, and 1 seasonal moving average term, also based on a yearly cycle, with a linear trend (drift) over time.

For the ARIMA with Dynamic Regression model, we took the best ARIMA model for each of the two cantons. We added two exogenous variables, first, a dummy variable for `Covid`, and the `exchange rate` between the Japanese Yen and Swiss Franc (JPY/CHF) was added to Luzern's ARIMA.

TSLM

For TSLM, a model that included the exchange rate between the Japanese Yen and Swiss Franc (JPY/CHF) and a Covid dummy was created for Luzern. The model was fitted using the following formula:

$$value = JPY + covid + trend + season$$

Model Selection

The model selection process involved evaluating three key metrics listed below for each model. The chosen models demonstrated the lowest metric values and exhibited the most understandable and visual forecast.

1. `AIC (Akaike Information Criterion)`: The AIC is a measure of the relative quality of a statistical model for a given set of data. It balances the complexity of the model against how well the model fits the data, penalizing excessive complexity to prevent overfitting.
2. `MAE (Mean Absolute Error)`: MAE quantifies the average magnitude of the errors in a set of predictions, without considering their direction. It's a straightforward metric that provides a quick indication of predictive accuracy at the individual level.
3. `MASE (Mean Absolute Scaled Error)`: MASE measures the accuracy of forecasts relative to a naive benchmark prediction, typically the naïve seasonal model. This scaling makes it an excellent choice for comparing forecast performance across different data scales and is particularly useful when dealing with seasonal variations.

Vaud Metrics

For Vaud, the ARIMA with Dynamic Regression variable of a Covid dummy was selected as the best model based on the AIC, MAE, and MASE metrics.

Model	AIC	MAE	MASE
ETS Vaud	5335.510	5339.696	0.4594194
Arima Vaud	4504.753	5065.484	0.4358266
COVID Arima Vaud	4396.572	4386.575	0.3774144

Luzern Metrics

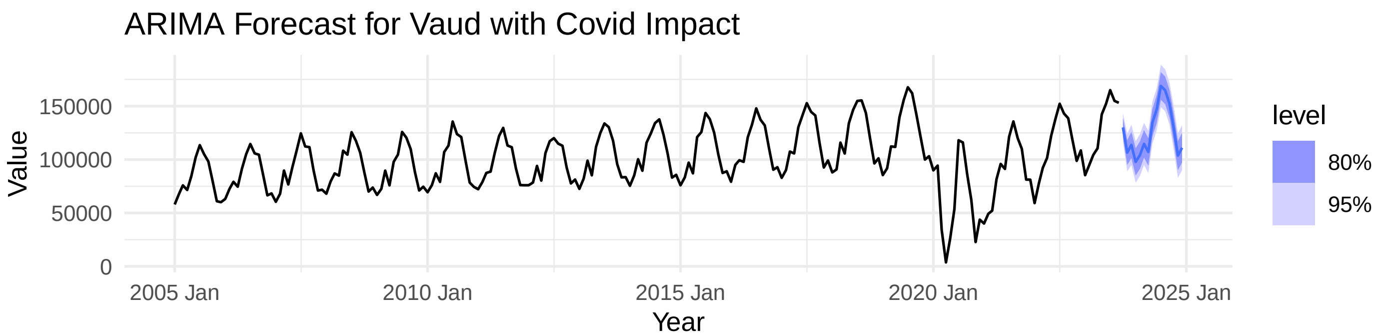
For Luzern, the results are not so clear, the ARIMA with Dynamic Regression variable of a Covid dummy and the exchange rate between the Japanese Yen and Swiss Franc (JPY/CHF) has the lowest AIC. However, the ARIMA with Dynamic Regression variable of a Covid dummy has the lowest MAE and MASE.

Model	AIC	MAE	MASE
ETS Luzern	3572.375	187.4152	0.8772346
Arima Luzern	2929.193	149.0430	0.6976257
JPY Arima Luzern	2931.172	149.2661	0.6986701
JPY COVID Arima Luzern	2927.563	150.2202	0.7031361
TSLM Luzern	3518.821	416.2821	1.9484921

Results

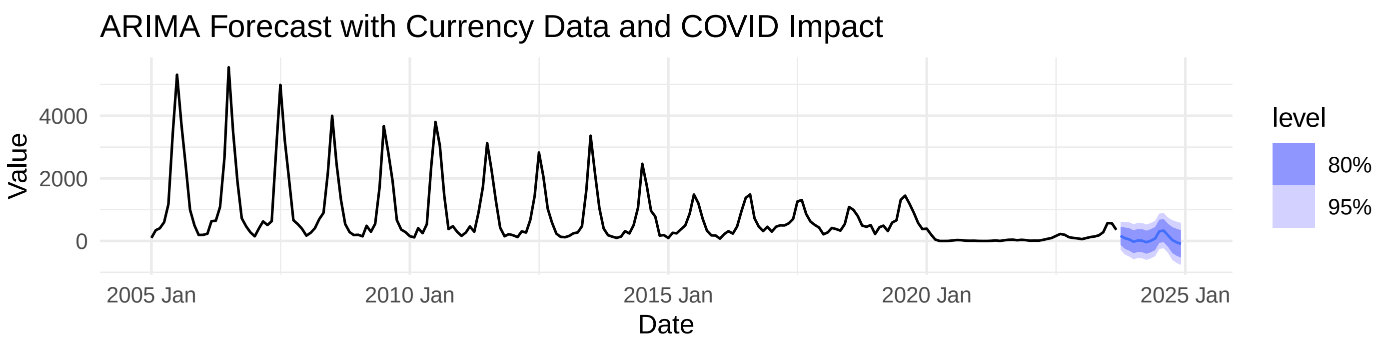
Vaud

The ARIMA with Dynamic Regression model for Vaud forecasts a steady increase in overnight stays, with a strong recovery since the impact of Covid. The model captures the seasonal fluctuations and the overall upward trend in tourist visits, providing valuable insights for tourism management in Vaud.



Luzern

The ARIMA with Dynamic Regression model, incorporating the dummy variable for COVID-19 and the exchange rate between the Japanese Yen and Swiss Franc, achieved the lowest AIC for predicting overnight stays by Japanese travelers in Lucerne. We chose this model, despite similar results across metrics, as using the exchange rate could enhance long-term forecast accuracy. The model forecasts a continuing downward trend. While the forecast displays negative values, these should be interpreted cautiously and considered as zero, indicating that the model predicts no overnight stays from Japanese tourists in Lucerne for certain months.



Note : From this forecast, negative values should be interpreted as zero, indicating that the model predicts no overnight stays from Japanese tourists in Lucerne for those months.

Conclusion

The results of our analysis present a mixed outlook. For tourism in Vaud, the trend appears positive, with forecasts indicating a steady recovery to pre-COVID levels. This suggests a strong rebound in tourist activity, aligning with the observed upward trend in overnight stays. On the other hand, the forecast for Japanese tourists in Lucerne shows a declining trend. This downturn, exacerbated by the significant appreciation of the Swiss Franc against the Japanese Yen, suggests that travel to Switzerland has become more expensive for Japanese tourists, likely contributing to the predicted decrease in overnight stays.

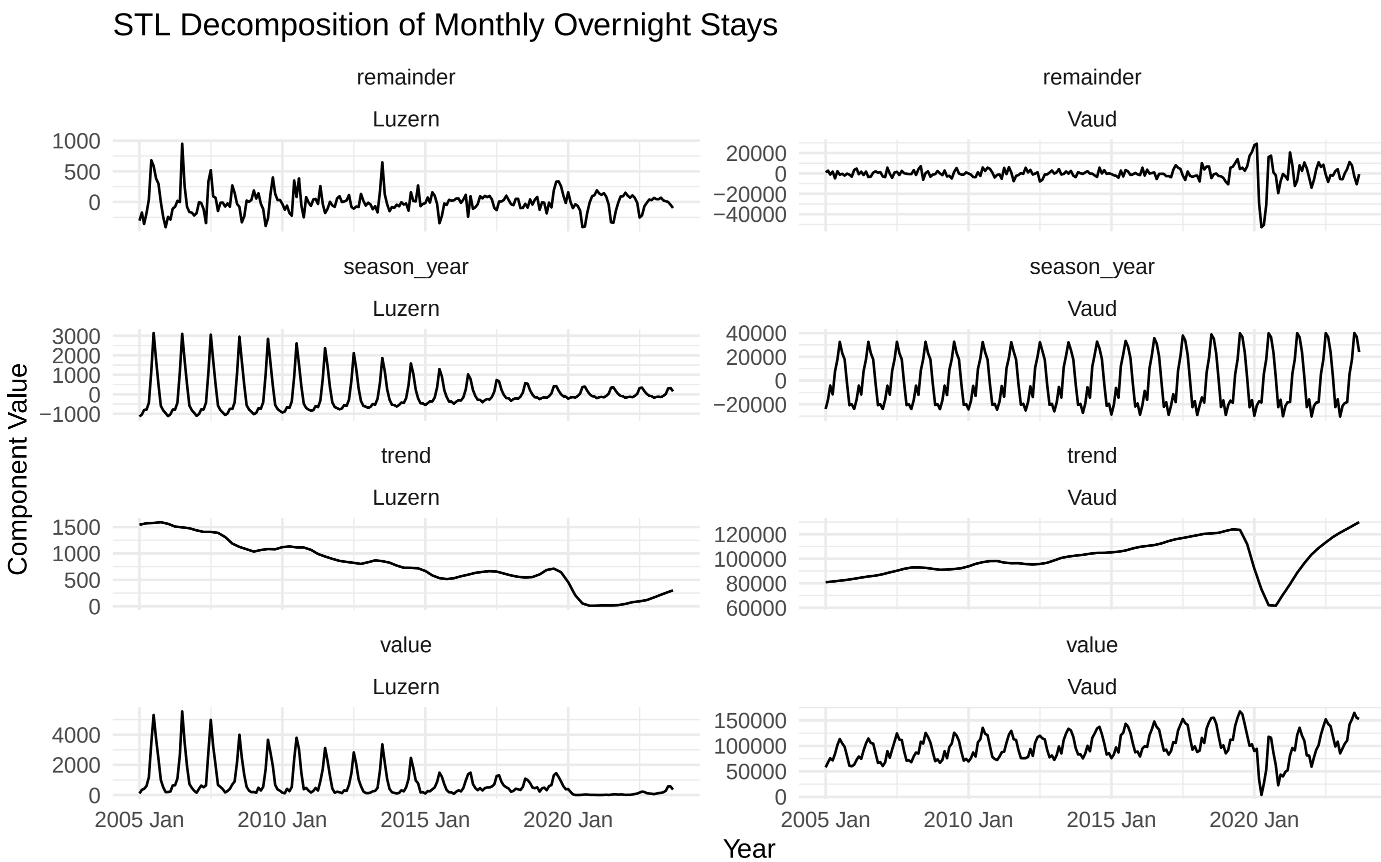
Our findings also underscore the value of incorporating external factors into our models. The ARIMA with Dynamic Regression model, enhanced with a COVID-19 dummy variable for Vaud and both the COVID-19 dummy and exchange rate for Lucerne, consistently outperformed other models. This highlights the importance of external variables in improving forecast accuracy. Future work could further enhance these models by including additional external variables such as weather conditions, public holidays, and other relevant factors. As demonstrated, integrating these variables could significantly improve the quality of forecasts, providing more reliable insights for tourism management and strategic planning.

Appendix

Appendix 1 - Sample Data

Herkunftsland	Kanton	Monat	Jahr	value
Indien	Obwalden	August	2022	518
Belarus	Aargau	Juni	2012	3
Hongkong	St. Gallen	April	2010	8
Schweiz	Neuchâtel	November	2009	4718
Frankreich	Basel-Stadt	Mai	2011	2410

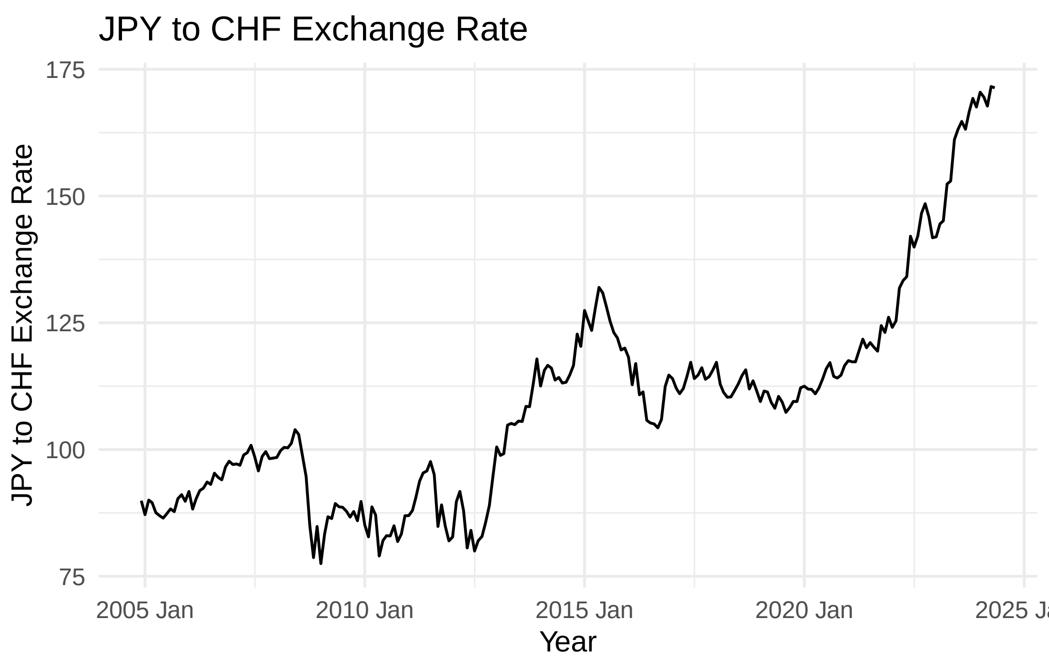
Appendix 2 - STL Decomposition of Luzern and Vaud



Appendix 3 - Exchange Rate JPY vs CHF

Links

- Data Source - chf/jpy
- Code Repository



JPY to CHF Exchange Rate
175
150
125
100
75
JPY to CHF Exchange Rate
2005 Jan
2010 Jan
2015 Jan
2020 Jan
2025 Ja
Year