Thai Tourism
Time Series Forecast

Presented By
Vicky Fan, Jane Han, Jenny Jiang, Ali Mansour

Agenda

- Executive Summary
- Data Overview and Preparation
- Characteristics of the Time Series
- Model Fitting
- Model Selection and Evaluation
- Conclusion and Future Work

OUR COMPANY

Forecasts is an analytics consultancy firm specializing in tourism data analysis and forecast as well as investigating variables associated with variability in tourism patterns and providing high quality recommendations to improve tourism.

Our Team

Vicky Fan

Jane Han

Jenny Jiang

Ali Mansour

Executive Summary

Executive Summary

Thai government is interested in investigating the trends and factors associated with tourism in Thailand.

As a proof of concept, FORECASTS is asked to consider the data on tourists arriving from the United States between the Years of 2010 and 2016.

Amongst others, FORECASTS will investigate seasonal and economic variables associated with the number of tourists arriving from the U.S.A. overtime.

FORECASTS will ,thereafter, explore various models to forecast tourism trends from U.S.A to Thailand and provide recommendations regarding ideal time windows during which the Thai government need focus its resources on accommodating tourists from the U.S.A.

Data Overview and Preparation

Data Overview and Preparation

Data Overview

Thailand Tourism Data
source: data.world

Currency Data
source: exchangerates.org.uk

Temperature Data
source: aws-observations.climate

Data Preparation

Original Data:

- Structured csv file.
- Data is recorded monthly.
- 4452 rows, 5 columns.

Processed Thailand Tourism Data:

- Only tourism data for the United States are used.
- Jan 2010 to Dec 2015 (Train).
- Jan 2016 to Dec 2016 (Test / Forecast Horizon).

Processed Currency & Temperature Data

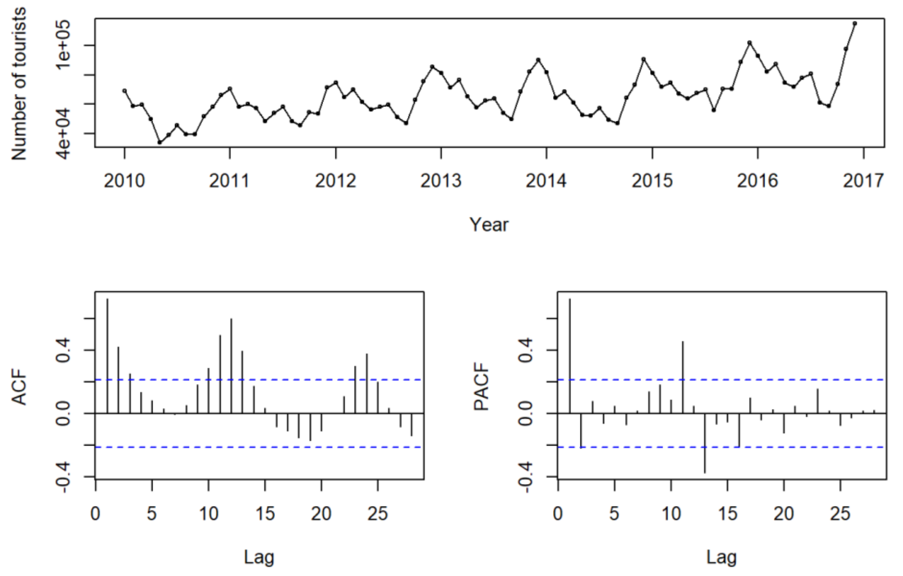
- Monthly average exchange rate of 1 USD vs. Thai Baht is used.
- Monthly average temperature in Bangkok is used.
- Jan 2010 to Dec 2015 (Train).
- Jan 2016 to Dec 2016 (Test / Forecast Horizon).

Characteristics of the Time Series

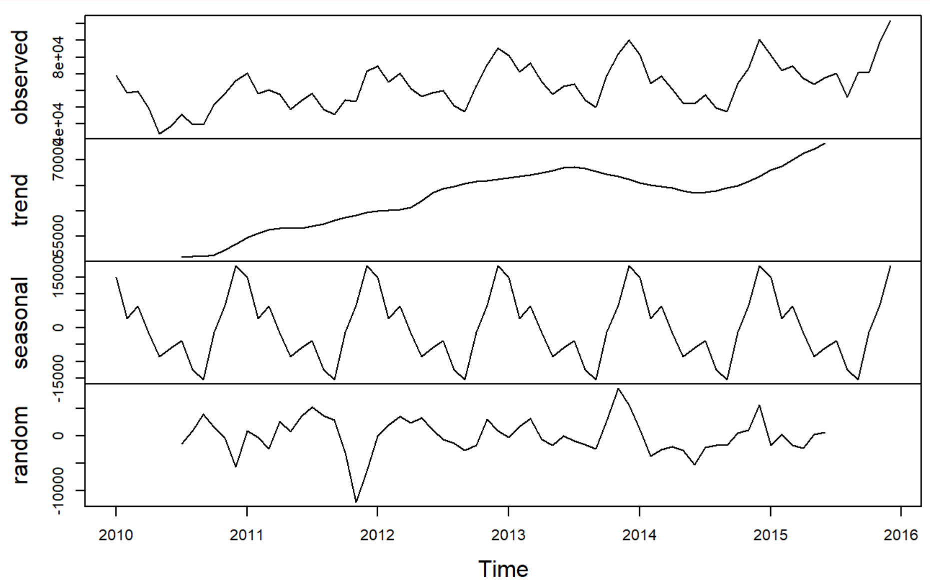
Characteristics of the Main Data

- Clear upward trend
- Annual seasonality
- Inconsistent variance over time
- KPSS test: Data is nonstationary

Number of Tourists Visiting Thailand By Month

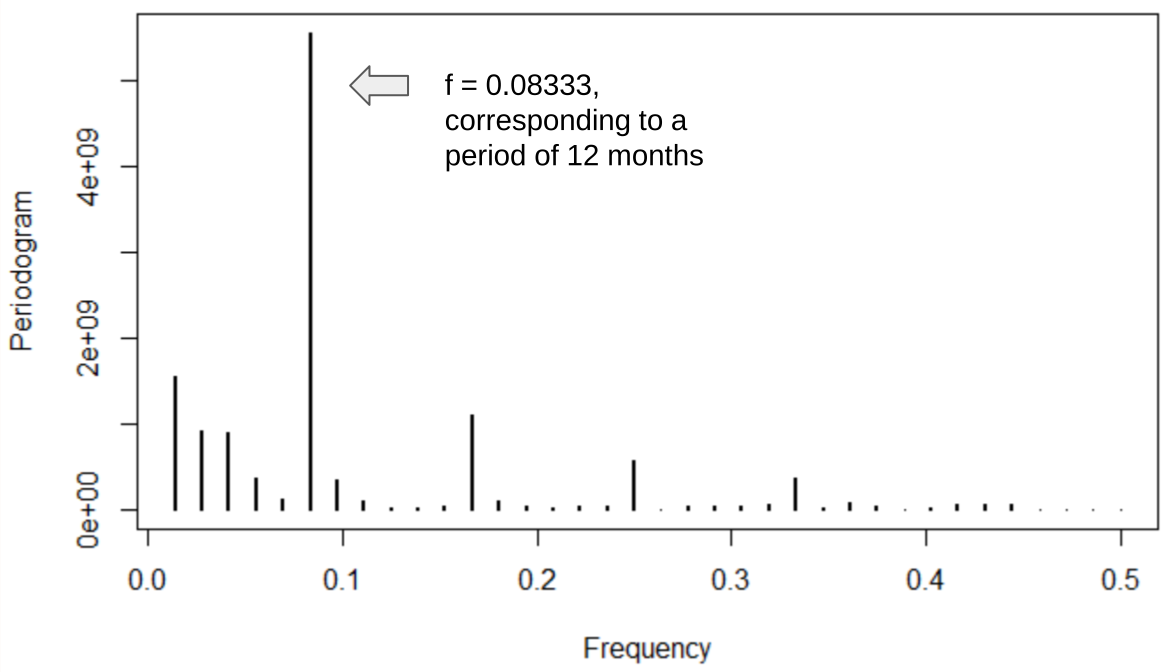


Decomposition of Additive Time Series



Characteristics of the Main Data

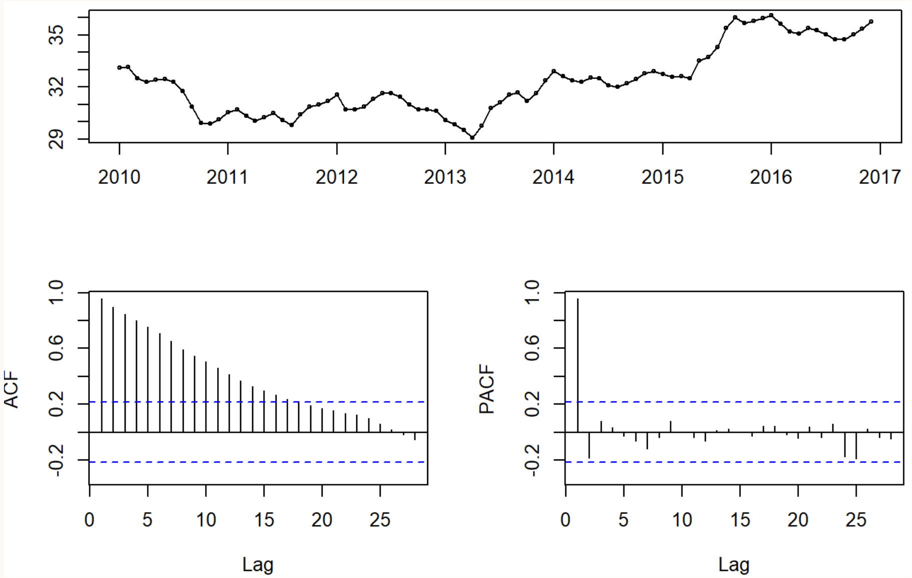
Periodogram shows annual seasonality is the dominant seasonal component



Characteristics of the Additional Data

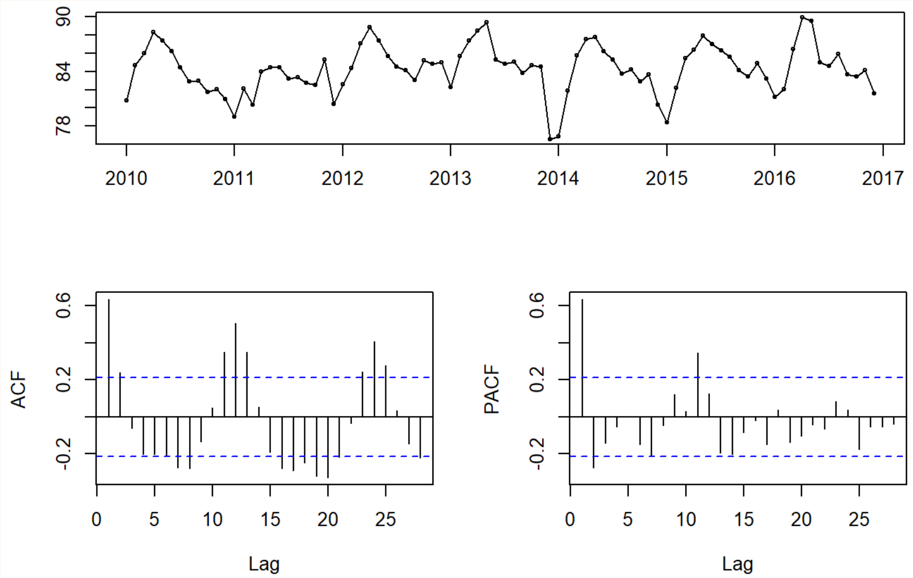
Monthly Average Exchange Rate – USD vs. Thai Bhat:

- Clear upward trend
- No clear seasonal pattern
- KPSS test: Data is nonstationary
- ACF decreases slowly



Monthly Average Temperature in Thailand (Bangkok)

- No clear trend
- Annually seasonal pattern
- KPSS test: Data is stationary

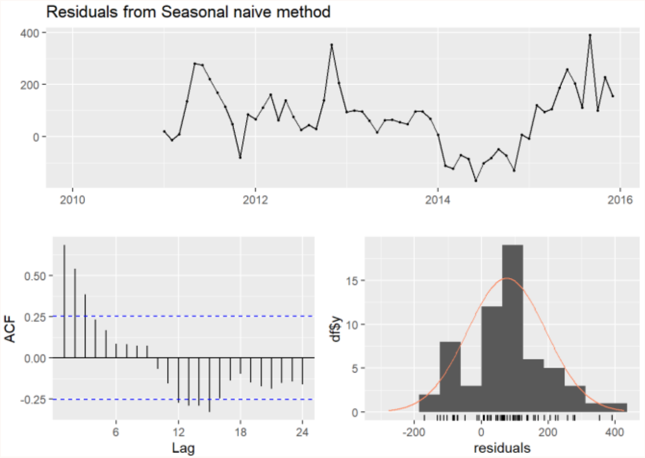


Model Fitting

Seasonal Naive

Model Phrase Outline

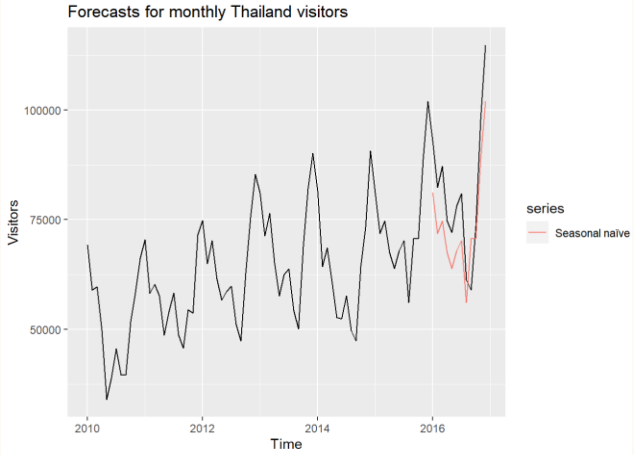
- Since the dominant seasonal component is annual seasonality, snaive model is applied and used as the baseline model.



Ljung-Box Test:
Residuals are not independent (p-value: 1.602e-12)

Results on Test Data

Model: Seasonal Naive
RMSE: 9841.895
MAE: 9394.833



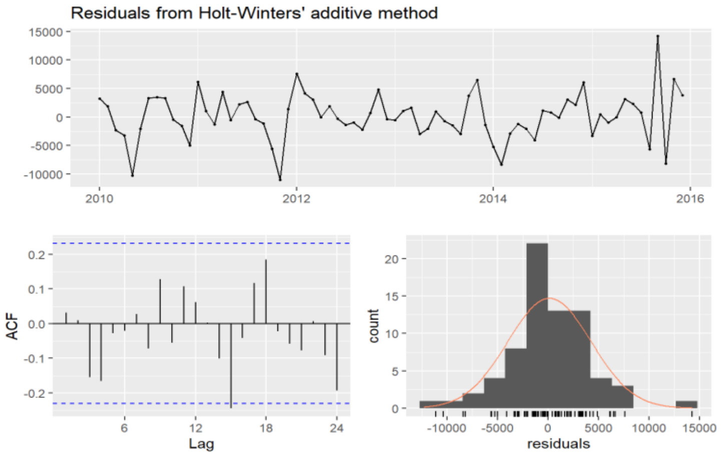
Holt-Winters

Model Phrase Outline

- Seasonality was modeled as Additive after observing that Multiplicative yielded a higher AICc
- The resulting smoothing parameters are:

  α (level) = 0.6036 β (trend) = 0.0001

  γ (seasonality) = 0.0001



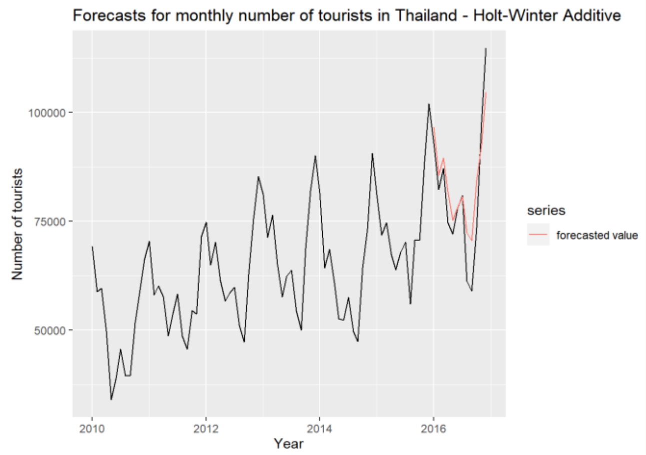
Ljung-Box Test: Residuals are not independent (p-value: 0.000265)

Results on Test Data

Model: Holt-Winters with Additive

RMSE: 7077.21

MAE: 5762.498

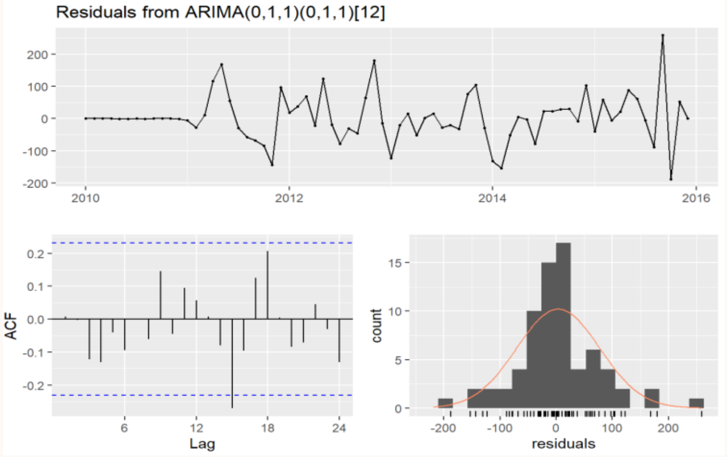


SARIMA

Model Phrase Outline

Model	AICc	BIC
ARIMA(1,0,0)(2,1,0)[12] with drift	706.1	715.46
ARIMA(0,1,1)(0,1,1)[12]	696.77	702.57

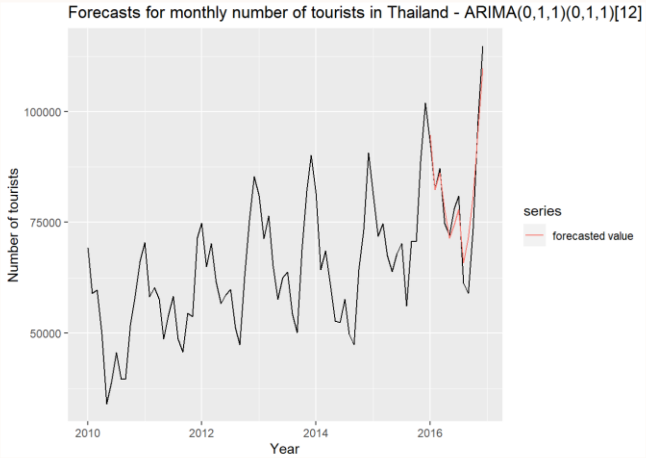
Box-Cox λ: 0.627



Ljung-Box Test: Residuals are independent (p-value: 0.8406)

Results on Test Data

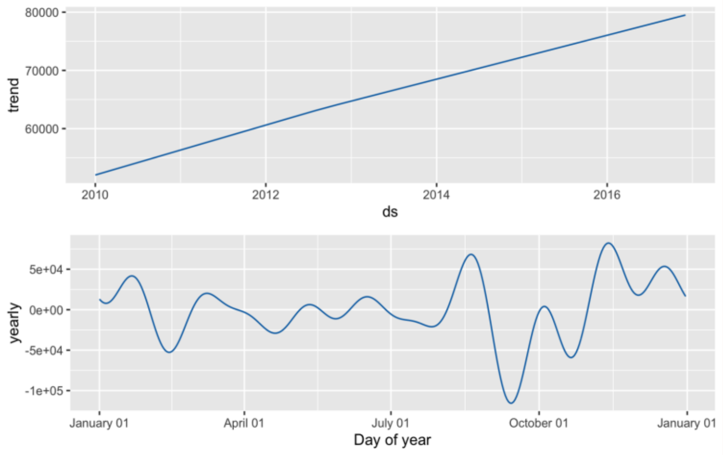
Model: ARIMA(0,1,1)(0,1,1)[12]
RMSE: 4995.538
MAE: 3749.619



Prophet

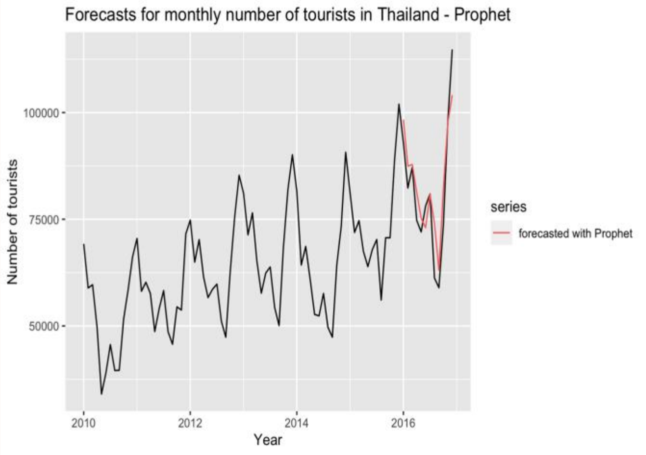
Model Phrase Outline

- Annual seasonality was modeled
- Our data is monthly so we are unable to model using holidays



Results on Test Data

Model: Prophet
RMSE: 6458.374
MAE: 5280.61



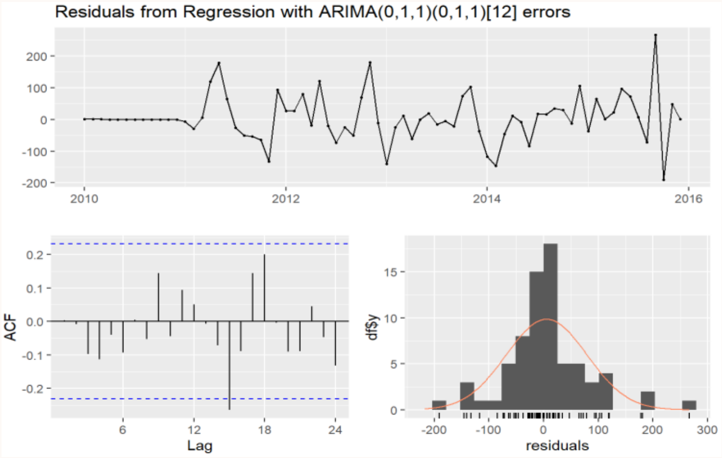
Regression w/ ARIMA errors

Model Phrase Outline

	ma1	sma1	Avg. exchange rate	Avg. temp.
Coeff.	-0.36**	-0.46**	-11.26	-2.61

**: 0.05 significant level

Box-Cox lambda: 0.627



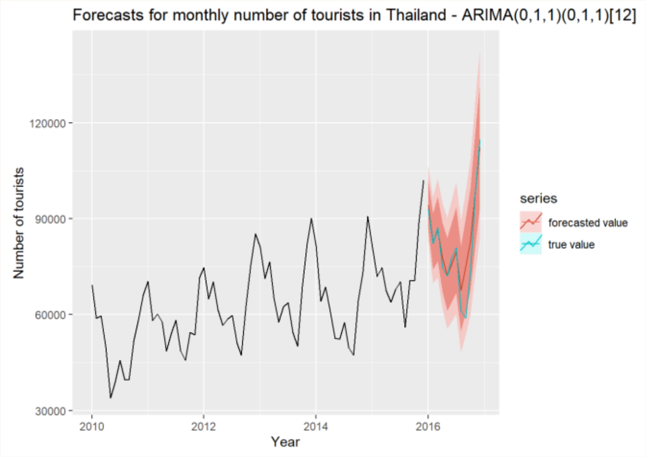
Ljung-Box Tes: Residuals are independent (p-value: 0.8022)

Results on Test Data

Model: Regression with ARIMA(0,1,1)(0,1,1)[12] errors

RMSE: 5814.869

MAE: 3595.783



Vector Autoregression Model (VAR)

Model Phrase Outline (1/3)

- Apply first order differencing (d = 1, D = 1) and Box-Cox transformation to stablize the variance and convert the data to stationary time series.
- Use VARselect to choose P orders:

AIC(n)	HQ(n)	SC(n)	FPE(n)
10	1	1	2

- First, we fit VAR(1) and run a serial.test. We find that there is no autocorrelation in the residuals of VAR(1)

VAR(1)

```
 Portmanteau Test (asymptotic)

data:  Residuals of VAR object var.fit.1
Chi-squared = 76.657, df = 81, p-value = 0.616
```

- So, we choose VAR(1) to conduct further diagnosis

1. Coefficients of VAR(1)

$tourists

tourists.l1	average.exchange.rate.l1	average.temperature.l1	const
0.642264***	-0.066997	0.062205	0.002347*

$average.exchange.rate

tourists.l1	average.exchange.rate.l1	average.temperature.l1	const
0.1791822	0.9447248***	-0.1161593	0.0006439

$average.temperature

tourists.l1	average.exchange.rate.l1	average.temperature.l1	const
0.1815577	0.0613772	0.5631571***	-0.0008766

Vector Autoregression Model (VAR)

Model Phrase Outline (2/3)

2. Diagnosis of VAR(1)

- Serial.test

```
 Portmanteau Test (asymptotic)

data:  Residuals of VAR object var.fit.1
Chi-squared = 76.657, df = 81, p-value = 0.616
```

There is no autocorrelation in the residuals.

- Arch.test

```
 ARCH (multivariate)

data:  Residuals of VAR object var.fit.1
Chi-squared = 282, df = 432, p-value = 1
```

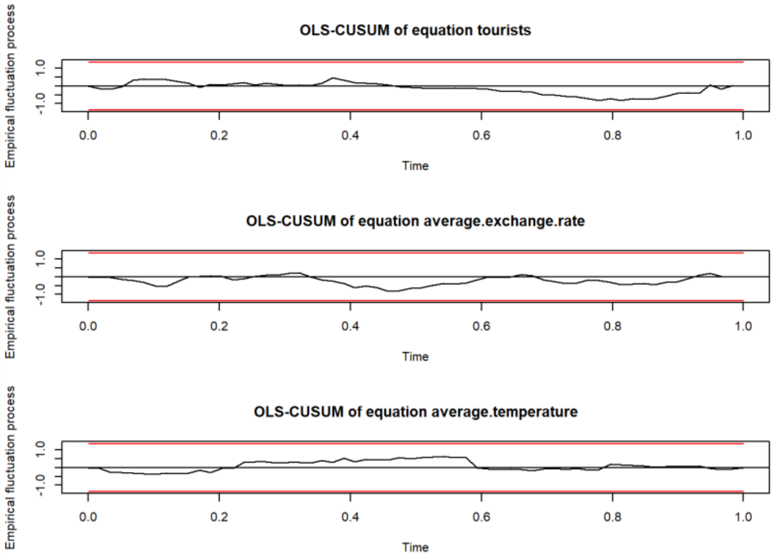
There is no heteroscedasticity in the residuals.

- Normality.test

  - JB Test: p-value < 2.2e-16
  - Skewness: p-value = 3.481e-05
  - Kurtosis: p-value < 2.2e-16

  Residuals are not normally distributed

- Structural break test



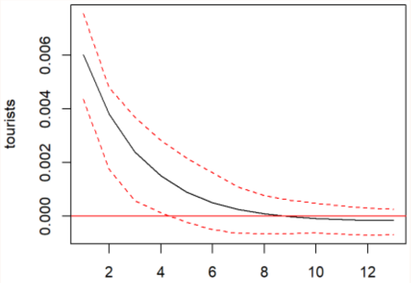
No structural break in the residuals

Vector Autoregression Model (VAR)

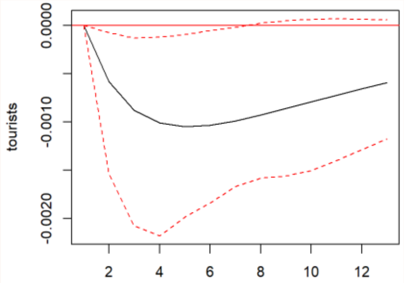
Model Phrase Outline (3/3)

- Impulse response test

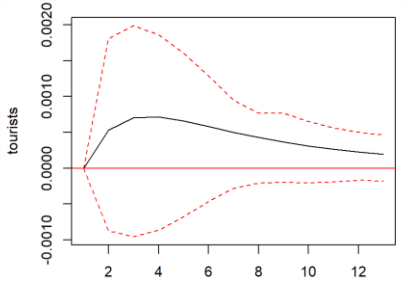
Impulse Response from tourists



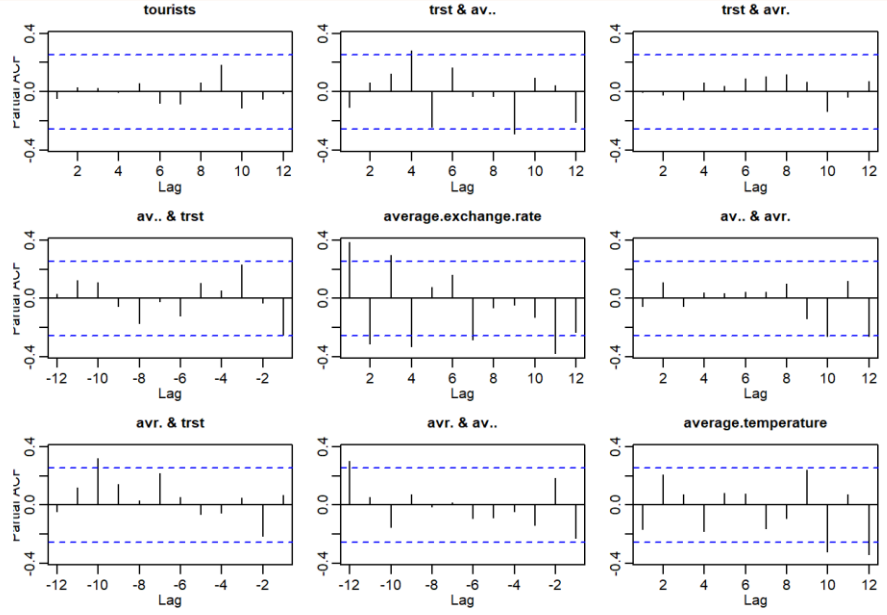
Impulse Response from average.exchange.rate



Impulse Response from average.temperature



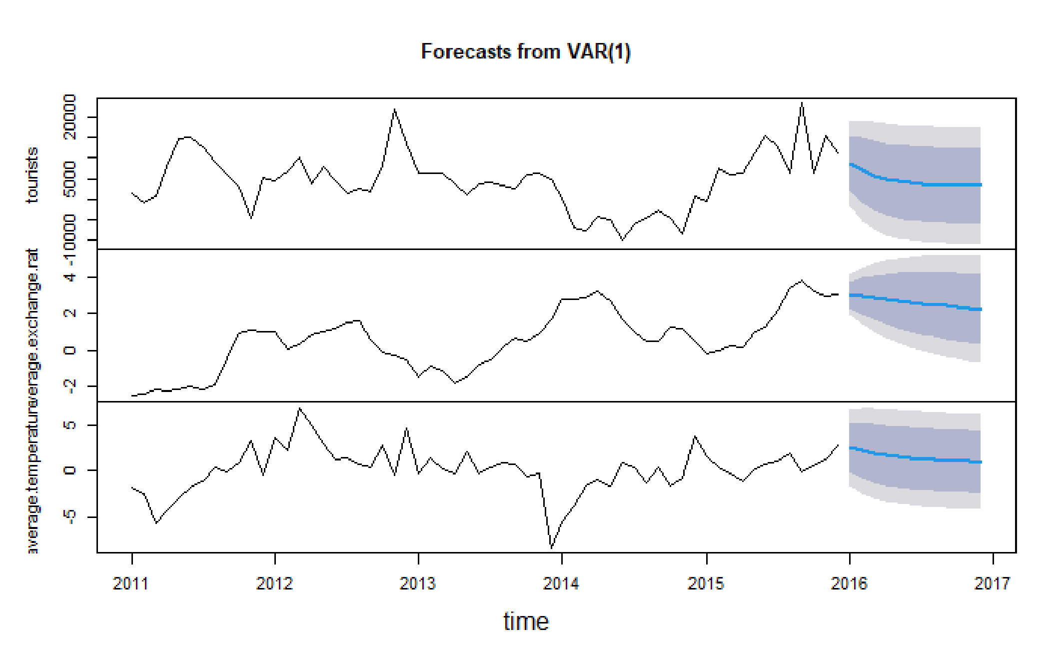
3. Residuals of VAR(1)



Vector Autoregression Model (VAR)

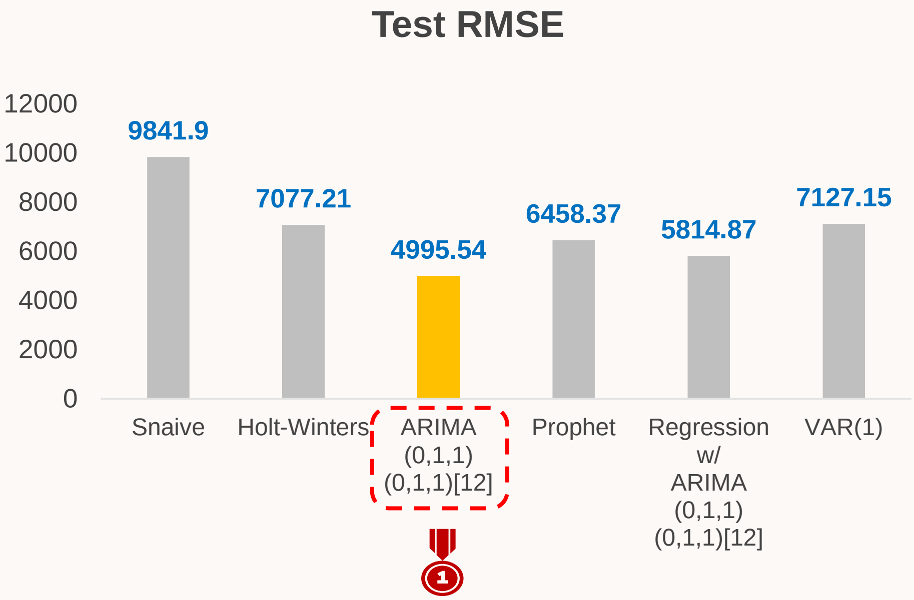
Results on Test Data

Model: VAR(1) RMSE: 7127.135 MAE: 8121.93



Model Selection and Evaluation

Model Selection

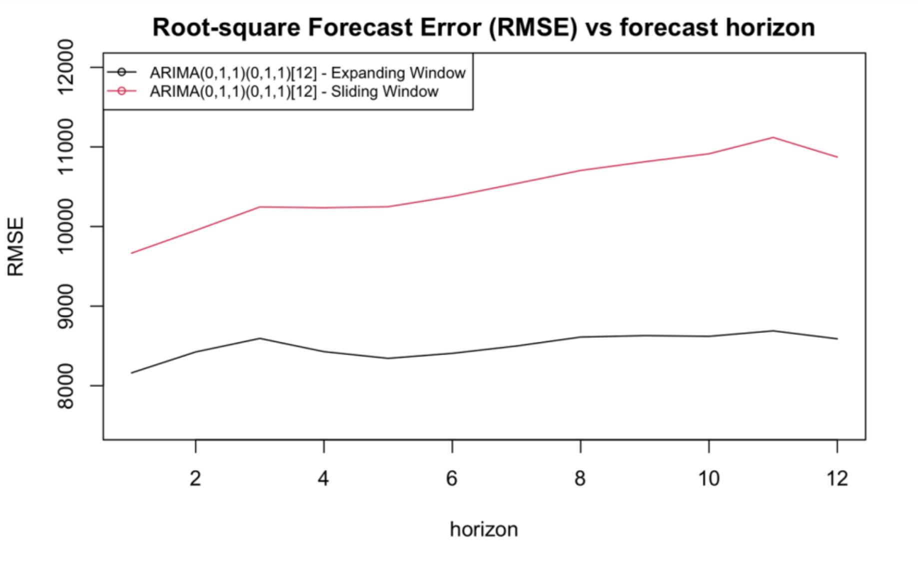


- Since no outliers exists in our dataset, RMSE is selected as the metric to evaluate model performance.

- Surprisingly, ARIMA(0,1,1)(0,1,1)[12] is the best model.

- According to results of Regression w/ ARIMA(0,1,1)(0,1,1)[12] and VAR(1), it's clear that average exchange rate and average temperature data are not good features to forecast the tourism data.
  - Granger causality test verifies this conclusion

Chosen Model Evaluation



- Apply cross-validation on the train dataset
- Chosen model is pretty stable
- The expanding window performs better with lower RMSE

Conclusion & Future Work

Conclusion and Future Work

Conclusion

- Trends of U.S. tourists to Thailand exhibit seasonality; while spikes exits in most quarters, the fourth quarter of the year sees the highest counts typically followed by a down-trend for the first three quarters of the following year.
- We recommend an increase in resource availability and targeted feature attractions (catered for tourists from the U.S.) during the fourth quarter of each year

Future Work

- Augment original tourism data with more data to further model stability
- Investigate additional data sets that may be used to develop more accurate models
- Investigate advanced time series models:
  - TBATS model
  - Prophet for multivariate analysis
  - Long-Short Term Memory model
- Follow up on the effect of suggested interventions and their effect on the number of U.S. tourists during and off peak season

Thank You

Q & A

Welcome for any question and feedback :)

Reference

- https://otexts.com/fpp2/VAR.html;
- https://medium.com/@swetha994r/time-series-forecasting-using-vector-auto-regressive-var-model-1ecd4a564ee0;
- https://medium.com/@cdabakoglu/time-series-forecasting-arima-lstm-prophet-with-python-e73a750a9887
- https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/
- https://towardsdatascience.com/vector-autoregressive-for-forecasting-time-series-a60e6f168c70
- https://medium.com/@swetha994r/time-series-forecasting-using-vector-auto-regressive-var-model-1ecd4a564ee0
- https://faculty.washington.edu/ezivot/econ584/notes/varModels.pdf
- https://stats.stackexchange.com/questions/47495/whats-a-stationary-var
- https://stats.stackexchange.com/questions/261876/var-for-non-stationary-series-using-r
- https://stats.stackexchange.com/questions/148994/var-or-vecm-for-a-mix-of-stationary-and-nonstationary-variables/149263#149263
- https://towardsdatascience.com/granger-causality-and-vector-auto-regressive-model-for-time-series-forecasting-3226a64889a6
- https://towardsdatascience.com/vector-autoregressions-vector-error-correction-multivariate-model-a69daf6ab618
- https://davegiles.blogspot.com/2011/04/testing-for-granger-causality.html

Project Work Distribution

- Our Company & Executive Summary - Ali
- Data Sources and Data Preparation – Xiaoqin/Jenny
- Characteristics of the Time Series – Jenny/Ali
- Snaive – Jane
- Holt Winters + sARIMA - Jenny
- Prophet - Jane
- Regression with ARIMA errors - Ali
- VARs - Xiaoqin
- Model Selection and Evaluation – Jane/Jenny
- Conclusions and Future Work – Xiaoqin