Project Outline

Sabrina Koshedub-Colaco

June 2025

1 Introduction

In December 2024, Donald Trump referred to then president Justin Trudeau as "Governor Justin Trudeau of the Great State of Canada" (CBS 2025). In January of 2025, Trump first threatens "economic force" as a means of annexing Canada. In February and March, Trump announces and enacts 25% tariffs on all Canadian goods (ABC 2025). The execution of these tariffs has since been quite volatile, but what this means for Canadians is uncertainty, and animosity. The resulting shift in political sentiment has caused Canadians to boycott American goods and travel in an attempt to fight fire with fire (Marketplace 2025).

The objective of this project is to use airline booking and flight data as a proxy to quantify the effect of this political sentiment change through statistical modeling. I aim to develop an automated and scalable model that can be used to investigate changes in flight demand (passengers flown/to fly) in or between any locations, for any reason. This will not only be a way to quantify exogenous events, but as an informative tool for airlines themselves.

1.1 Question

Is there evidence to support that a drop in flight demand from Canada to Orlando is a result of the Trump administration in 2025? If so, what can be said about the nature, direction, and magnitude of this change?

1.2 Objectives

Without a baseline understanding of how flight demand between Canada and Orlando behaves under normal conditions, measuring the effect of a change in the conditions will be impossible. My first objective is to model and understand the underlying patterns (trends and cycles) of the relevant flight paths. Then, I shall be able to forecast how flight volumes may have developed in 2025 under normal conditions. Finally, I can compare my projections to the actual development of flights into 2025, and quantify the political impact on 2025 flight demand through intervention analysis.

2 Economic Theory

2.1 Keynesian Uncertainty

Theory: Keynes described “animal spirits” as psychological drivers of economic behavior. When confidence falls, consumption and investment fall—even if fundamentals are unchanged.(Akerlof and Shiller 2009)
Relevance: A drop in international travel to a country after a political event may not reflect price or income changes, but instead a sentiment-driven pullback.

2.2 Consumer Sentiment and Demand

Theory: Sentiment indexes (like the Michigan Consumer Sentiment Index (Michigan. Surveys of Consumers 2025)) are strong predictors of spending. Negative sentiment lowers consumers’ perceived wealth or future income expectations, decreasing current demand.(Carroll, Fuhrer, and Wilcox 1994)
Relevance: Negative political sentiment could dampen demand for discretionary travel.

2.3 Precautionary Saving and Reduced Consumption

Theory: When uncertainty increases, households increase savings to hedge against future risks, reducing current consumption.(Deaton 1991)
Relevance: In times of political instability, foreign tourists may avoid vacations to avoid perceived economic risks.

3 Readings

3.1 Intervention Analysis with Applications to Economic and Environmental Problems

Authors: George E. P. Box, George C. Tiao

Publication: Journal of the American Statistical Association, 1975 (Box and Tiao 1975)

Summary: The paper formalizes the intervention analysis technique, incorporating it within the ARIMA modeling structure. It distinguishes between different types of interventions (step, pulse, ramp) and provides a rigorous methodology for quantifying and testing the impact of these interventions on time series.

Relevance: This paper provides the theoretical and methodological basis for applying intervention analysis to assess how, in my case, political shocks may affect flight demand trends.

3.2 Using intervention time series analyses to assess the effects of imperfectly identifiable natural events: A general method and example

Authors: Stuart Gilmour, Louisa Degenhardt, Wayne Hall, Carolyn Day

Publication: BMC Medical Research Methodology, 2006 (Gilmour et al. 2006)

Summary: The study outlines an approach for modeling the effects of interventions that are only partially identifiable, incorporating techniques for estimating the timing and form of the intervention effect. The method is applied using heroin overdose data in response to a suspected reduction in heroin supply.

Relevance: This paper is valuable for modeling the impact of political interventions on flight demand when the intervention's effects are diffusing or gradual rather than sharply defined.

4 Theoretical Model / Empirical Specifications

4.1 Understanding the Baseline

I expect to see both trends and seasonality in my data. Trends are long-term, consistent directions in the data (for example: a general upward trend in flight volume after lifting COVID-19 travel bans). Seasonality refers to repeating patterns within a specific period (for example: a yearly spike in flight demand around Christmas time).

4.1.1 X-13ARIMA-SEATS

X-13ARIMA-SEATS is a seasonal adjustment method developed by the United States Census Bureau. This method decomposes a time series into three components: trend, seasonality, and irregularity.

$$Y_t = T_t + S_t + I_t \tag{1}$$

Where Y_t is the observed value at time t, T_t is the trend component, S_t is the seasonal component, and I_t is the irregular component. The X-13ARIMA-SEATS model uses regARIMA modeling (a regression model with ARIMA errors) to pre-adjust the series and identify and correct for outliers. This modeling also allows me to factor in moving holidays. Isolating the trend and seasonal components provides a clear view of long-term demand patterns and helps identify structural breaks or policy effects, which will be further analyzed in intervention models.

4.1.2 ACF/PACF

Autocorrelation Function (ACF):

$$\rho_k = \frac{\mathrm{Cov}(Y_t, Y_{t-k})}{\mathrm{Var}(Y_t)} \tag{2}$$

Where ρ_k is the autocorrelation at lag k, $\mathrm{Cov}(Y_t, Y_{t-k})$ is the covariance between Y_t and its lagged value, and $\mathrm{Var}(Y_t)$ is the variance of the series. The ACF measures how correlated a time series is with its own past values (lags). An ACF at lag 12, for example, may suggest the presence of annual seasonality in monthly data. ACF does not distinguish whether the correlation at lag k is a direct effect of lag k, or an indirect effect of intermediate lags.

Partial Autocorrelation Function (PACF):

$$\phi_{kk} = \mathrm{Corr}\left(Y_t - \hat{Y}_t^{(k-1)},\, Y_{t-k} - \hat{Y}_{t-k}^{(k-1)}\right) \tag{3}$$

Where ϕ_{kk} is the partial autocorrelation at lag k, $\hat{Y}_t^{(k-1)}$ is a linear projection of Y_t on $Y_{t-1}, \ldots, Y_{t-k+1}$, and $\hat{Y}_{t-k}^{(k-1)}$is a linear projection of Y_{t-k} on $Y_{t-1}, \ldots, Y_{t-k+1}$. The PACF measures the direct correlation between Y_t and Y_{t-k} after removing the linear influence of all intermediate lags $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k+1}$. For example, the PACF at lag 3 shows how much of Y_t is explained specifically by Y_{t-3}, not due to the cumulative influence of Y_{t-1} and Y_{t-2}. This makes PACF a useful tool for identifying the number of autoregressive (AR) terms to include in my model.

4.2 Forecasting Normal Conditions

Considering the trends and seasonality found in the above methods, I shall project what flight demand may have been in 2025 had there been no shocks.

4.2.1 SARIMA

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is a common and powerful framework for modeling time series data that exhibit both trend and seasonality. The general form of a SARIMA model is written as:

$$\textbf{SARIMA}(p,\, d,\, q)(P,\, D,\, Q)_s$$

$$\Phi_P(B^s)\,\phi_p(B)\,(1-B)^d\,(1-B^s)^D Y_t = \Theta_Q(B^s)\,\theta_q(B)\,\varepsilon_t \tag{4}$$

Backshift Operator (B): The backshift operator is defined as $BY_t = Y_{t-1}$, and allows efficient representation of lagged terms. Higher powers of B correspond to further lags, for example, $B^2Y_t = Y_{t-2}$.

Differencing for Stationarity: By adjusting for nonstationarity, the model ensures that forecasts are based on stable and interpretable relationships in the data.

The terms $(1-B)^d$ and $(1-B^s)^D$ perform non-seasonal and seasonal differencing, respectively. $(1-B)^d$ removes trends by differencing the series d times. For example, if $d=1$, the transformation becomes $Y_t' = Y_t - Y_{t-1}$. $(1-B^s)^D$ removes seasonal effects by subtracting the value from s periods ago. For monthly data with annual seasonality, $s=12$ and $D=1$ implies $Y_t' = Y_t - Y_{t-12}$.

Autoregressive (AR) Components: These terms model the influence of past values of the series.

Moving Average (MA) Components: These terms model the influence of past forecast errors (residuals). If recent predictions were too high or too low, the model adjusts the current forecast based on those prior mistakes.

Error Term (ε_t): This component captures the part of the series that is not explained by trend, seasonality, or autocorrelation.

Interpretation: The SARIMA model first transforms the original series into a stationary series using differencing. Then it models the remaining structure using both autoregressive (AR) and moving average (MA) terms at both regular and seasonal lags.

4.3 Interrupted Time Series/Intervention Analysis

Intervention analysis is a statistical technique specifically designed to assess the impact of exogenous shocks on time series data.

4.3.1 SARIMAX

$$Y_t = \mu + \beta X_t + \text{SARIMA terms} + \varepsilon_t \tag{5}$$

Where μ is the intercept, X_t is an exogenous variable (such as an intervention dummy, in my case, 0 before Jan 2025, and 1 after), β is the coefficient for X_t, and all other terms are as previously defined

SARIMAX extends the SARIMA framework by incorporating exogenous variables that are believed to influence the time series being modeled. In the context of this project, the exogenous variable represents the re-election of Donald Trump and the resulting shift in U.S.–Canada relations. With this, I shall be able to estimate how flight demand deviates from expected seasonal patterns in response to this external shock. Depending on the nature of the intervention, this variable may take one of several forms: a step function, which equals 0 prior to the intervention and 1 thereafter, modeling a sustained level change, a pulse function, which equals 1 only in the period of the intervention, representing a short-lived shock, or a ramp function, which increases/decreases incrementally over time following the intervention, modeling a gradual impact. I will investigate which form is best for my intervention with visualizations and media research.

The estimated coefficient of the intervention variable provides a measure of the magnitude and direction of the political event's impact on flight demand (number of passengers flown/to fly). By comparing the SARIMAX model's fitted values to observed data and to my SARIMA projection assuming no intervention, I can formally assess whether and to what extent the event disrupted normal travel patterns.

5 Data

Both data sources include a passenger count, which is the variable I shall be using to quantify demand.

5.1 T100

The T100 data comes from the Bureau of Transportation Statistics website. The variables I chose to include are: '`fltmth`' field which records the flight month with dates like 2021-04-01, '`opr_al`' which identifies the airline operating the flight using its IATA code (such as DL for Delta Airlines), the origin and destination of each flight are captured by '`apt_fm`' and '`apt_to`' (airport codes), and '`ctry_fm`' and '`ctry_to`' (ISO country codes), directional flight paths are specified using '`dir_apt_pair`' and '`dir_ctry_pair`', with non-directional equivalents being '`ndir_apt_pair`' and '`ndir_ctry_pair`'. Each row also includes numeric fields: '`flt_cnt`' (number of flights), '`seat_cnt`' (total seats), and '`psgr_cnt`' (total passengers).

5.2 Booking data

I was lucky to get access to airline booking data, which is collected by third party websites for congruency in booking purposes. This dataset consists of the flight month, the snapshot date of when the data was collected, the directional country pair of travel, point of sale country, and the total number of booked passengers as of the snapshot date. The real value of this dataset lies in the fact that it gives real insights to future demand patterns through bookings.

References

ABC (2025). *Trump keeps talking about making Canada the 51st state. Is he serious?* ABC News. URL: https://abcnews.go.com/Politics/trump-talking-making-canada-51st-state/story?id=119767909&utm.

Akerlof, George A. and Robert J. Shiller (2009). *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism.* Princeton, NJ: Princeton University Press.

Box, George E. P. and George C. Tiao (1975). “Intervention Analysis with Applications to Economic and Environmental Problems”. In: *Journal of the American Statistical Association* 70.349, pp. 70–79. DOI: 10.1080/01621459.1975.10480264.

Carroll, Christopher D., Jeffrey C. Fuhrer, and David W. Wilcox (1994). “Does Consumer Sentiment Forecast Household Spending? If So, Why?” In: *American Economic Review* 84.5, pp. 1397–1408.

CBS (2025). *Trump trolls Justin Trudeau as "governor" of the "great state" of Canada.* CBS News. URL: https://www.cbsnews.com/news/trump-trolls-justin-trudeau-governor-canada/?utm.

Deaton, Angus (1991). “Saving and Liquidity Constraints”. In: *Econometrica* 59.5, pp. 1221–1248. DOI: 10.2307/2938366.

Gilmour, Stuart et al. (2006). “Using intervention time series analyses to assess the effects of imperfectly identifiable natural events: A general method and example”. In: *BMC Medical Research Methodology* 6.16. DOI: 10.1186/1471-2288-

6-16. URL: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-6-16.

Marketplace (2025). *As trade war drags on, Canadians avoid U.S. products — and travel*. Marketplace. URL: https://www.marketplace.org/story/2025/05/19/canadians-respond-to-trade-war-with-us-product-boycotts.

Michigan. Surveys of Consumers, University of (2025). *Surveys of Consumers*. URL: https://www.sca.isr.umich.edu/.