The Proceedings of Statistics Canada Symposium 2018
Combine to Conquer: Innovations in the Use of Multiple Sources of Data

A cross-border traveller's tale: Do I stay or do I go?

Tanja Armenski, Zdenek Patak, and Steve Matthews[1]

Abstract

This article integrates weather and traffic data to better understand the impact of weather on cross-border traffic, and to develop modelling techniques that may be well suited to producing accurate forecasts. Traffic data are collected by the Canadian Border Service Agency (CBSA) and produced by Statistics Canada as Frontier Counts (FC), an administrative data on all international arrivals to Canada. Weather data such as temperature, rainfall and snowfall were obtained from Environment and Climate Change Canada. ARIMA models including weather and holiday-related regressors confirm that significant weather events such as snowfall and freezing rain cause reductions in the number of commuters crossing the border. Statutory holidays also represent significant departures from the norm and should be taken into account when modelling and forecasting cross-border traffic flows.

Keywords**:** Data Integration; ARIMA; Weather; Cross-Border Traffic.

1. Introduction

Cross-border traffic is a key economic indicator of international trade and tourism. United States residents entering Canada by car comprise the largest highest share of all international arrivals. Weather – particularly severe weather – is often seen as a factor influencing the volume of traffic. Long stretches of unusually low or high temperatures, or significant snowfall may impact an individual's decision to take a trip across the border. Weather data, at least in part, may explain the variation in car volume in traffic time series.

The objective of this study is two-fold. First, this study incorporates data from two distinct sources, (i) traffic data from the Canada Border Services Agency (CBSA), compiled by Statistics Canada into the Frontier Counts (FC) program, and (ii) weather data from Environment and Climate Change Canada. Second, to provide a more systematic approach incorporating weather into estimation of traffic volumes across the Canada–US border. The main results show that major snowfall and freezing rain have a significant impact on cross-border traffic. Each weather phenomenon causes a reduction in the volume of car traffic across the border. Cross-border traffic also decreases significantly on some Canadian holidays while it increases on others.

1.1 Literature review

The relationship between weather and traffic is widely recognized. Two consistent findings are noted. First, there is an association between traffic flows and severe weather conditions. Datla et al. (2013), demonstrate that the association between traffic flows in Alberta and the cold and snowy weather varies with respect to the severity of weather

[1] Tanja Armenski, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Canada, K1A0T6; Zdenek Patak, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Canada, K1A0T6; Steve Matthews, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Canada, K1A0T6

conditions. Warm temperature in the summer months drives increases in traffic volumes while cold weather with snowfall results in traffic declines during winter months. Roh et al. (2016) show that there is a higher reduction for passenger car volumes than truck volumes on cold days in Alberta. The reductions in car and truck traffic volumes intensify with higher amounts of snowfall due to cold and snow interaction.

Second, the impact of adverse weather on traffic movements varied by day of the week. Datla et al. (2013), show that volume patterns are more affected by cold on weekends compared with weekdays. There is now a growing interest in modelling the impact of weather on traffic systems in real time, or nowcasting. This modelling approach was initially applied to weather conditions to increase the spatial accuracy of weather conditions and its impact, but then was adopted for traffic systems to increase the prediction accuracy of traffic movements in real time (Lin, et al., 2015). With the availability of high frequency (daily) data released just ten days after the end of the reference month or six weeks in advance of Frontier Counts, the Modern Tourism Statistics Program is also a step-closer to traffic data collection in real time. The paper is divided as follows: Section 2 describes traffic and weather data; Section 3 explains various estimation methods and results, followed by concluding remarks in Section 4.

2. Data sources and variables

2.1 Traffic data

The Canada Border Services Agency (CBSA) collects information on travelers entering the country. The Integrated Primary Inspection Line (IPIL) is an automated system that records daily number of vehicles entering Canada via land ports. All major CBSA ports have incorporated IPIL to replace the paper border entry form in an effort to modernize their statistical systems.

The IPIL system automatically generates information in a digital format, decreasing data processing requirements and increasing the timeliness of traffic data. The IPIL is a major component of the administrative Frontier Counts (FC) data as it covers all major land ports of Canada's international car traffic and is used to help create leading indicator of cross-border traveller volumes, capturing travellers entering and returning to Canada by car. Outbound traffic, Canadian passenger vehicles returning from the USA, is analyzed in this paper. Tourism partners and provincial users can now observe daily changes in the cross-border vehicular traffic. Integrating alternate data to support the statistical needs is one of the key modernization initiatives of Modern Tourism Statistics Program.

2.2 Land port of interest

The land port selected for this study is Cornwall[2], a land border crossing located in Southeast Ontario that connects Highway 138 in Ontario with Highway 37 in New York State. The crossing is associated with the Seaway International Bridge, which crosses the two channels on each side of Cornwall Island located in the St. Lawrence River. Traffic data is collected daily for the period from March 1, 2016, to April 30, 2018. Daily weather data are available from Environment Canada for the corresponding period. Average weather statistics along with the corresponding cross-border traffic counts for the period under study are shown in Table 2.2.1-1.

Table 2.2.1 -1
Climate and traffic data for period under study (March 1, 2016 to April 30, 2018)

[2] All Ontario land ports were studied. Cornwall was selected to illustrate the types of analyses performed on the data.

Average max temp °C	Average annual rainfall mm	Average annual snowfall cm	Average number of cars per day	Average number of cars weekday	Average number of cars weekend
12.3	798.1	234.3	1,407	1,400	1,424

When the average number of cars with Canadian licence plates returning from the US is broken down by the day of the week, Friday shows the heaviest traffic with Tuesday being the lightest. The difference between the two days is approximately 150 cars.

2.3 Weather data

Weather data were obtained from Environment and Climate Change Canada. Daily weather observations such as maximum and minimum temperatures, rainfall and snowfall are obtained from individual weather stations throughout Canada. Geographic coordinates were used to select the weather stations with the closest proximity to the land entry ports on the US-Canada border in Ontario. Weather variables used in this analysis include (a) maximum temperature (degrees Celsius); (b) indicators of significant snowfall (equal to one when snowfall exceeded 5cm and zero otherwise) on weekdays and weekends; (c) indicator of freezing rain (equal to 1 when rainfall was greater than 0 and air temperature was below 2° C[3] and zero otherwise); and (d) indicators of previous day snow accumulation on weekdays and weekends.

2.4 Holiday effects

Initial investigations were conducted to test the impact of holidays on cross-border traffic despite weather conditions. Significantly higher traffic was observed during statutory holidays such as Victoria Day, Civic Holiday, Canada Day, Labour Day, Easter Monday and Thanksgiving Day and lower traffic during Christmas Day and New Year's Eve. Two dummy variables were used to represent the effect of holidays on traffic volumes. The first dummy variable takes on a value of one when the holiday has a positive impact on traffic and zero otherwise. The second dummy variable takes on a value of one when the holiday has a negative impact on traffic and zero otherwise. The grouping of the holidays may not be ideal, but the restricted length of the series was not conducive to constructing regressors for individual holidays to take advantage of their full predictive ability. As more data become available, the holiday regressors will be refined.

3. Methods and results

3.1 Identifying significant lags (or frequencies)

When analyzing time series data, the first step is typically identifying the order of the non-seasonal and seasonal autoregressive and moving average components of the series by analyzing the autocorrelation (ACF) and partial autocorrelation (PACF) functions at various lags (e.g. weekly pattern represents a 7-day lag). To do so reliably, a series needs to be weakly stationary, i.e., the time series has a constant mean and autocovariance with respect to time (see Box and Jenkins, 1976). If a trend is present in the data, the stationarity criterion is violated. This is an important step as lack of stationarity may lead to erroneous conclusions. Mean stationarity can be verified using the Augmented Dickey-Fuller (ADF) test.

[3] For more information regarding formation norms, distribution and frequency of freezing rain in Ontario, please see the work of Cortinas (2000).

The ACF and PACF plots of the time series of cross-border traffic at Cornwall indicated seasonality at lags 7 (weekly) and 365 (annual). This supports the notion that the days of the week and the months of the year are not all equal. One would expect more travel on weekends and during the summer months. Other autoregressive or moving average parameters may be obtained by analyzing the ACF and PACF plots of the model residuals. A well-fitting model should have independent and approximately normally distributed errors, and ACF and PACF plots should have no significant lags. The next section describes the various models used in the analysis and the corresponding results in a hierarchical sequence designed to show how the residual mean squared error is reduced as the base model is being modified to include additional explanatory variables.

3.2 ARIMA model

The well-established ARIMA model can be used to explain variations in cross-border traffic. The base model starts with a pure seasonal ARIMA formulation with no regressors that can be written as:

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D y_t = \theta(\mathrm{B})\Theta(B^s)\varepsilon_t,\ d, D \geq 0 \tag{1a}$$

where $\phi(.)$ and $\Phi(.)$ specify the non-seasonal and seasonal autoregressive components of the model, and $\theta(.)$ and $\Theta(.)$ denote the non-seasonal and seasonal moving average components of the model. The backshift operator is denoted by B. The non-seasonal and seasonal differencing are denoted by d and D. The variable y_t is the value of the time series at time t, and ε_t are error terms satisfying $\varepsilon_t \sim N(0, \sigma^2)$, $Cov(\varepsilon_t, \varepsilon_s) = \begin{Bmatrix} 0, & \text{for } t \neq s \\ \sigma^2, & \text{for } t = s \end{Bmatrix}$.

The final base model was established to be $(1,1,1)(1,0,1)_7(0,0,1)_{365}$. This means that after the series has been first-differenced, AR(1), AR(7), MA(1), MA(7) and MA(365) components were included. The corresponding model can be expressed as

$$Y_t = (1+\phi_1)Y_{t-1} - \phi_1 Y_{t-2} + \Phi_7 Y_{t-7} - \Phi_7 Y_{t-8} + \theta_1 \varepsilon_{t-1} + \Theta_7 \varepsilon_{t-7} + \Theta_{365}\varepsilon_{t-365} + \varepsilon_t \tag{2a}$$

The resulting model residuals were independent and approximately normally distributed. However, the residual diagnostics indicated a heavy-tailed distribution and the presence of outliers that potentially imply model inadequacy. This is addressed in the next steps as the base model evolves.

Weather, in particular extreme cold temperature and large accumulations of snow and freezing rain, is assumed to influence people's decision to travel. To test this hypothesis, a number of weather related variables were added to the base ARIMA model. These included maximum daily temperature, an indicator of freezing rain, indicators of significant snowfall on weekdays and the weekend, and indicators of previous day significant snowfall on weekdays and the weekend.

The augmented model can be written as:

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D\left(y_t - \textstyle\sum_i \beta_{1i}x_{1it}\right) = \theta(\mathrm{B})\Theta(B^s)\varepsilon_t,\ d, D \geq 0 \tag{1b}$$

where x_{1it} is the value of the explanatory variable $1i$ at time t, β_{1i} is the regression coefficient associated with variable $1i$. The other parameters are identical to those of equation 1a.

The final augmented model incorporating weather regressors was $(1,1,1)(1,0,1)_7(0,0,1)_{365}$, or

$$Y_t = (1+\phi_1)(Y_{t-1} - \sum_i \beta_{1i} x_{1i,t-1}) - \phi_1(Y_{t-2} - \sum_i \beta_{1i} x_{1i,t-2}) + \Phi_7(Y_{t-7} - \sum_i \beta_{1i} x_{1i,t-7}) - \Phi_7(Y_{t-8} - \sum_i \beta_{1i} x_{1i,t-8}) + \theta_1 \varepsilon_{t-1} + \Theta_7 \varepsilon_{t-7} + \Theta_{365} \varepsilon_{t-365} + \varepsilon_t \quad (2b)$$

3.3 Results

The residual distribution was still heavy tailed as outliers had not been addressed at this point in the model evolution. All the weather variables were significant. The model estimates show that snowfall presents a major obstacle to cross-border travel. The impact of snow is greater on weekends when people may travel more for pleasure than for business and may be more likely to be able to modify their travel plans accordingly. Maximum temperature, although statistically significant, is not empirically relevant given that the number of cars crossing the border changes by marginal amounts (in the order of 3 cars per degree Celsius).

Adding weather variables to the base model decreased the heaviness of the tails of the residual distribution but a number of outliers persisted. Some large departures from the weekly pattern could be attributed to various Canadian holidays. The model incorporating a holiday regressor was identical to 2b with $\beta_{1i}x_{1it}$ being replaced by $\beta_{2i}x_{2it}$. By including two holiday indicator variables (one for positive impact of holidays and one for negative) regressors further decreased the heaviness of the tails of the residual distribution. The estimates for the weather variables are similar with the previous model, but now the two holiday regressors are showing a large impact on travel. “Positive” holidays (such as Victoria Day, Labour Day, etc.) increase traffic flow by 574 cars on average, whereas “negative” holidays (such as Christmas Day, Boxing Day, etc.) show an average decline of 421 vehicles relative to the well-established weekly pattern. These are significant departures from the norm and need to be taken into account when modelling and forecasting cross-border traffic flows.

3.4 Goodness of fit criteria

The goodness of fit of a statistical model describes how well it fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the fitted model. To assess the performance of the three models described above, variance of the model residuals and Akaike Information Criterion ($AIC = 2k - 2ln\hat{L}$, where k is the number of estimated parameters in the model and $\hat{L}$ is the maximum value of the likelihood function for the model) are computed. The model among the three candidates that minimizes the residual variance and the information loss is selected.

Table 3.4-1
Goodness of fit statistics

	Base ARIMA model	+ Weather regressors	+ Holiday regressors
Residual variance	18,087	15,401	11,339
AIC	9,999	9,881	9,644

The table shows that adding weather regressors to the base ARIMA model reduces the residual variance by 15%. A further 26% improvement can be realized by incorporating two holiday regressors. The numbers clearly indicate that adding pertinent regressors leads to significant gains in efficiency. As the model evolves and tracks the observed data more closely, it could be used to produce reliable and efficient forecasts, which is the ultimate goal of this modelling exercise.

4. Concluding Remarks

There are many factors that impact cross-border vehicular traffic. This study identifies two major ones, weather and holidays. Significant snowfall and freezing rain are weather events that cause a major reduction in the number of commuters crossing the border. This is not a surprising finding given that the corresponding road conditions become more hazardous and authorities encourage travelers to keep their driving to a minimum. The impact of adverse weather on traffic movements varied by day of the week with volume patterns being more affected by snowfall on weekends compared to weekdays. Weekend travellers might be more flexible in choosing whether to make a trip based on weather conditions at a given time, while travellers who travel regularly during weekdays, most likely for work or business, are less sensitive to the weather variations. A lagged snowfall effect on reduction of traffic was evident, indicating that when a large amount of snowfall occurs later in the day, road conditions the next day could still be treacherous enough to impact commuter travel plans.

The reduction in car traffic intensifies with the interaction of cold temperatures and rain. The study shows a negative impact of freezing rain on traffic volumes at the border crossings. Icy conditions make driving potentially hazardous and keep people off the road. The effect of temperature showed that the mean daily traffic volumes significantly increase with the warmer weather. The maximum temperature variable, although significant, is not really relevant given that cross-border traffic increases only for a few cars as temperatures reach the daily maximum. Various holidays on either sides of the border can also lead to either increased or reduced cross-border traffic. The largest increases in traffic volumes were recorded on Victoria Day. For many travellers it signifies the arrival of milder weather and possibly the start of summer travel season. On the negative side, Christmas Day stands out as a period of reduced travel.

The analysis and the corresponding output presented in this paper represent the first step toward understanding the data and identifying the models that provide the best fit and that could be used to forecast frontier counts depending on the weather, holidays and other pertinent regressors. This is the ultimate goal of the project, forecasting in general and possibly nowcasting in particular, as reliable forecasts could be used to manage resources at the Canadian land border crossings.

References

Box, G.E.P., and G.M. Jenkins (1976), *Time Series Analysis, Forecasting and Control (Revised Edition),* San Francisco: Holden-Day.

Cortinas, J. J. (2000), "A Climatology of Freezing Rain in the Great Lakes Region of North America", *Monthly Weather Review*, 128, pp. 3574- 3588.

Datla, S., S. Prasanta, H. Roh, and S. Sharma (2013), "A Comprehensive Analysis of the Association of Highway Traffic with Winter Weather Conditions", *Procedia - Social and Behavioral Sciences*, pp. 497 – 506.

De Livera, M. A., J. R. Hyndman, and D. R. Snyder (2011), "Forecasting time series with complex seasonal patterns using exponential smoothing", *Journal of the American Statistical Association*, 106:496, pp. 1513-1527.

Gelper, S., R. Fried, and C. Croux (2010), "Robust forecasting with exponential and Holt–Winters smoothing", *Journal of forecasting*, 29(3), pp. 285–300.

Lin, L., M. Ni, Q. He, J. Gao, and W. A. Sadek (2015), "Transportation Research Record: Journal of the Transportation Research Board", 2482, Transportation Research Board, Washington, D.C., pp. 82–89.

Roh, H., P. K. Sahu, S. Sharma, and S. Datla (2016), "Statistical Investigations of Snowfall and Temperature Interaction with Passenger Car and Truck Traffic on Primary Highways in Canada", *Journal of Cold Regions Engineering*, 30(2).