Time Series Analysis
Border Crossing

Rodrigo Meneses Orellana

27 de octubre de 2019

Introduction

An study of the behavior of border crossing data since January 1996 until March 2019 based on time series.

Data review

Taken from GP databases.

Information related with the border crossing between US-Mexico and US-Canada grouped by the most common ways.

In order to present a not exhaustive study based on time series, currently unnecessary information will be suppressed, leading to the next variables (the study could be expanded).

Preview:

Table 1: Database sample

year	month	Measure	Value
2019	Mar	Trucks	34447
2019	Mar	Rail Containers Full	428
2019	Mar	Trucks	81217
2019	Mar	Trains	62
2019	Mar	Personal Vehicle Passengers	16377
2019	Mar	Trucks	179

Data will be aggregated by year and month in order to form the time series. Aggregation leads to a stacked structure where each rectangle is an aggregation by Year/Month and Measure.

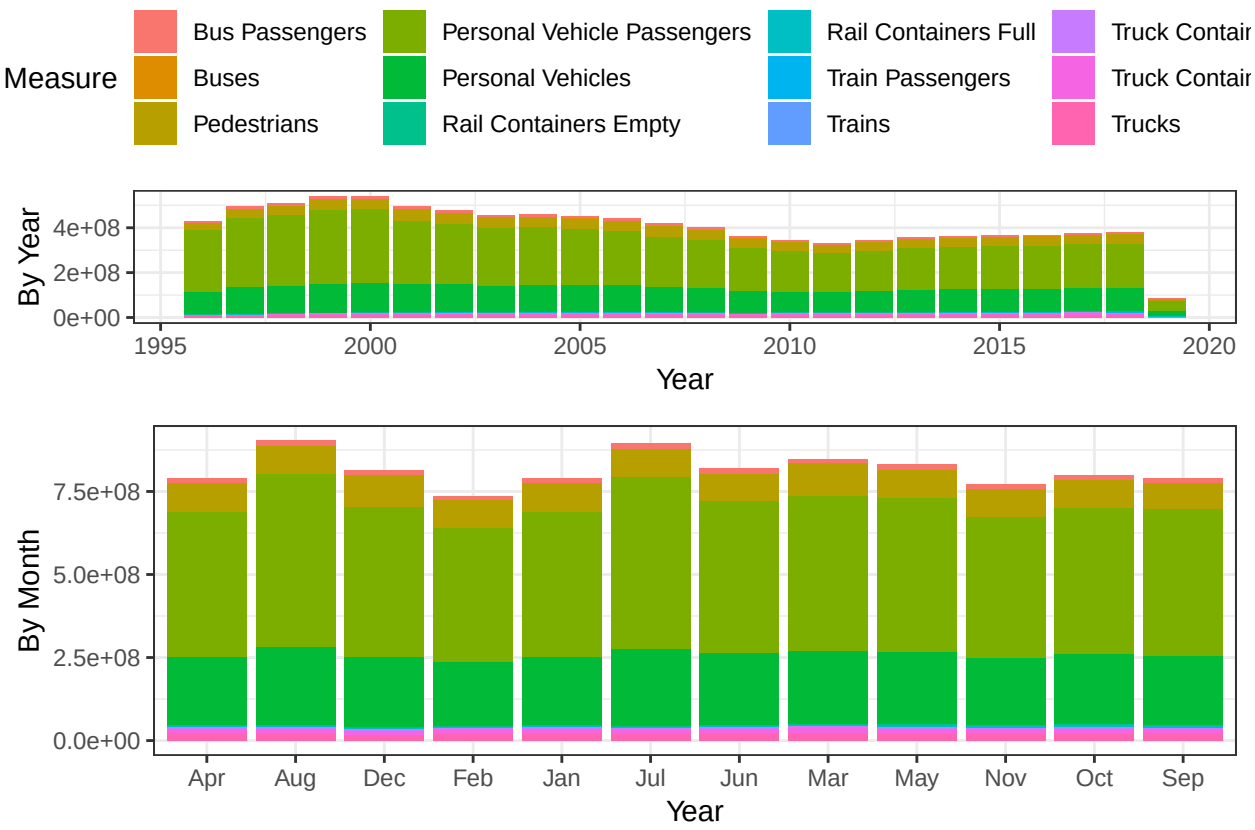

Figure 1: Data aggregation

Constructing and analyzing the Time Series

Aggregation by month and year leads to the time series data needed.

Then, it's easy to construct the time series.

Table 2: Time series data sample

year	month	value
1996	1	35690321
1997	1	38632051
1998	1	42047792
1999	1	43042565
2000	1	44639294
2001	1	43731679

Then, here's a visualization of the time series.

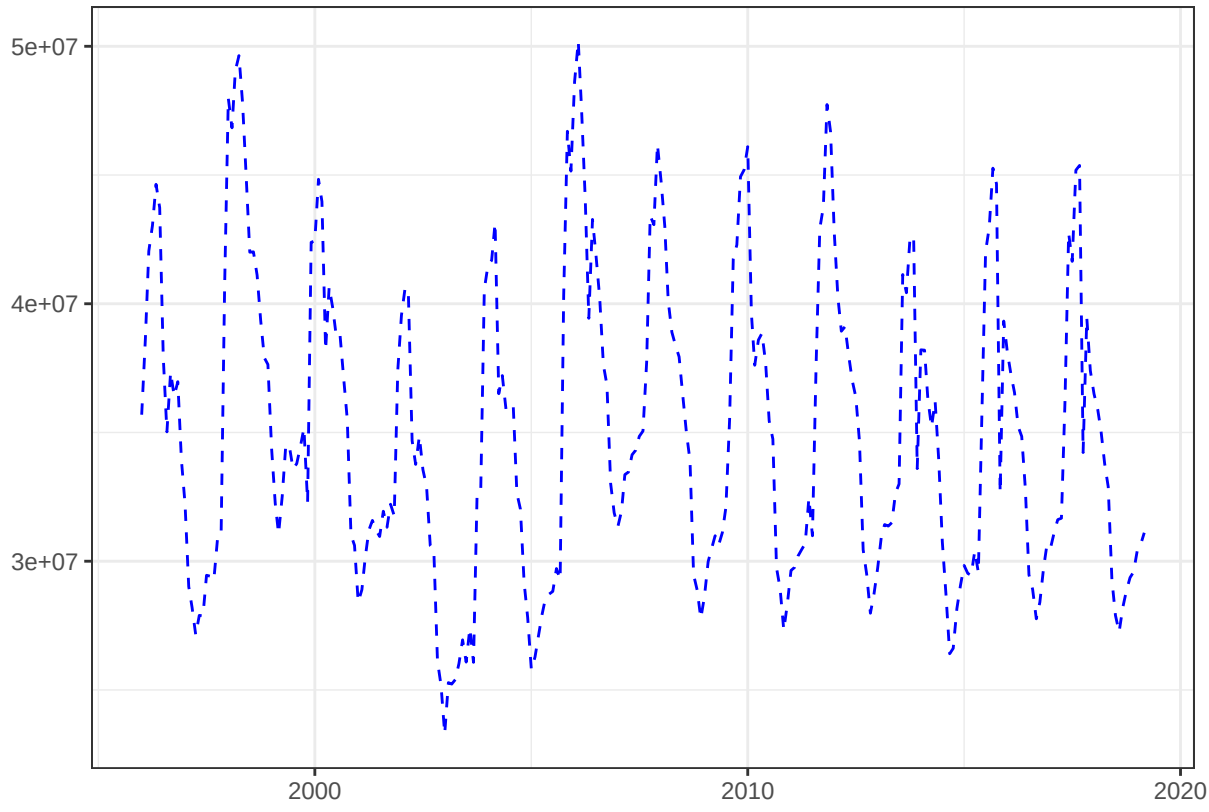

Figure 2: Time series

The time series seems to be stationary about the mean and variance. As a visual prove the autocorrelation function will be plotted.

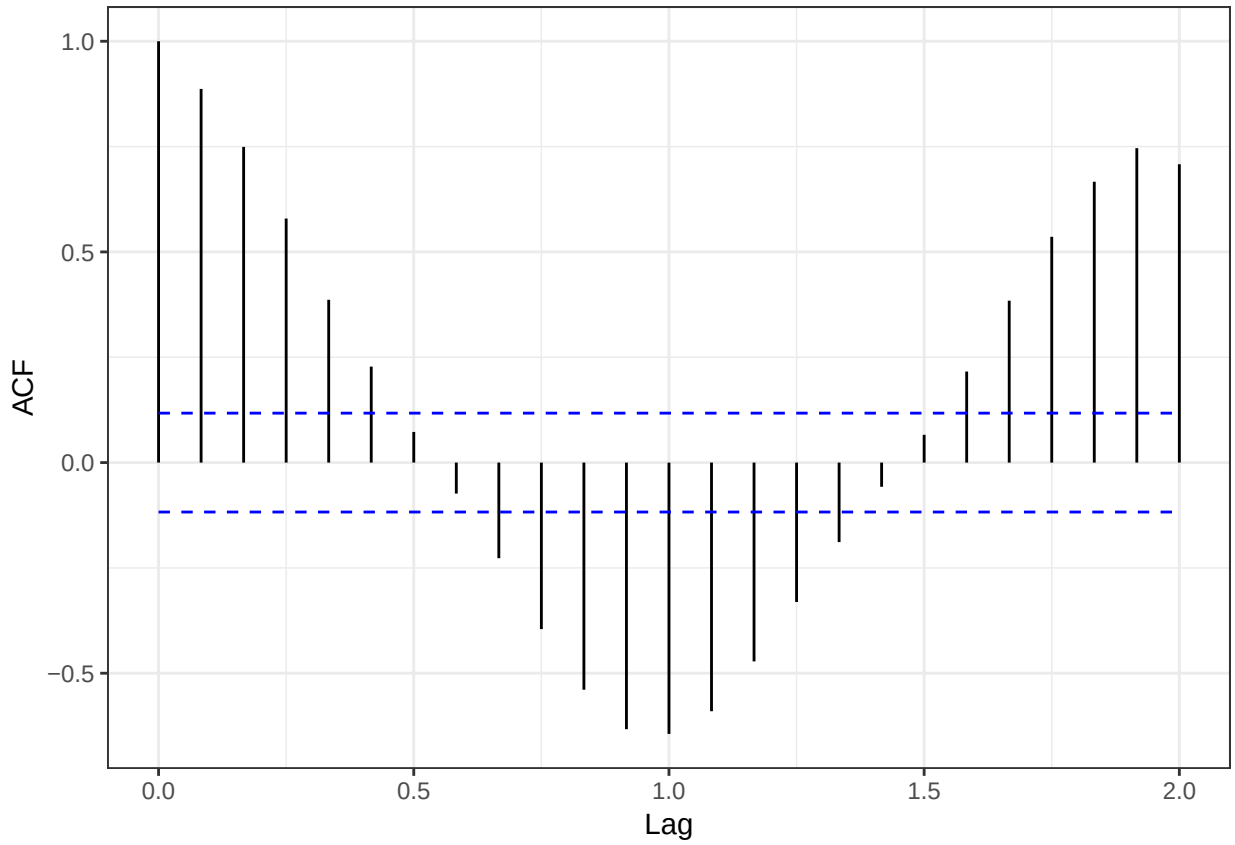

Figure 3: Autocorrelation function

Looking at the autocorrelation function is possible to say that the time series is stationary.

Next, each component of the time series will be studied.

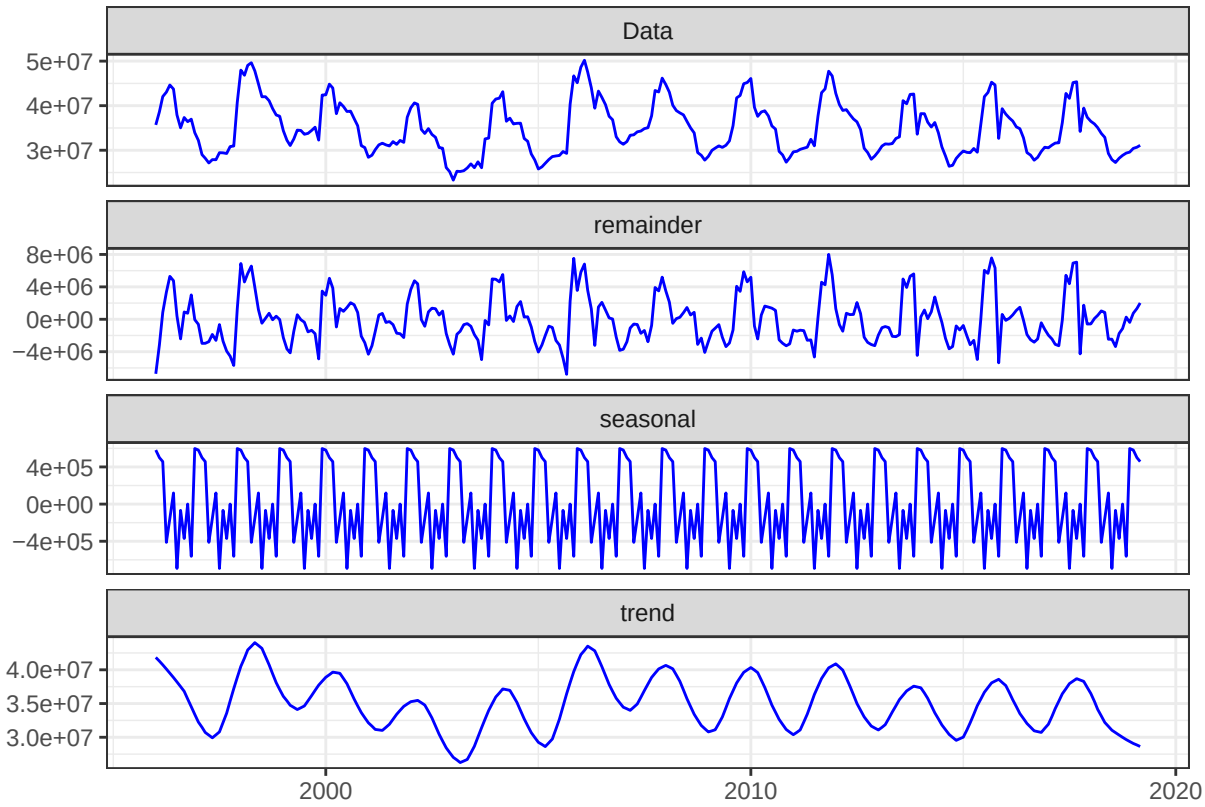

Figure 4: Component disaggregation

Given the component analysis, stationarity is fairly acceptable. Time series could be modeled by an ARIMA model. It's necessary to use a test to prove the stationarity. Then, ADF and KPSS tests will be applied where the null hypothesis is: **time series is non-stationary**.

Table 3: ADF and KPSS tests for non-stationarity

ADF	KPSS
0.01	0.1

Then, non-stationarity is rejected, so an ARIMA model will be fitted.

Residuals need to be independent (just white noise) and not to be normally distributed, therefore, is necessary to apply some tests.

Table 4: Residual independence tests

Box.Pierce	Ljung.Box
0.8803503	0.8797113

Then, **residuals aren't correlated**.

Table 5: Normal distribution tests for residuals

Jarque.Bera	Shapiro.Wilk
0	0

Then, **residuals aren't normally distributed**.

Finally, the model is ready to be deployed, some predictions will be made.

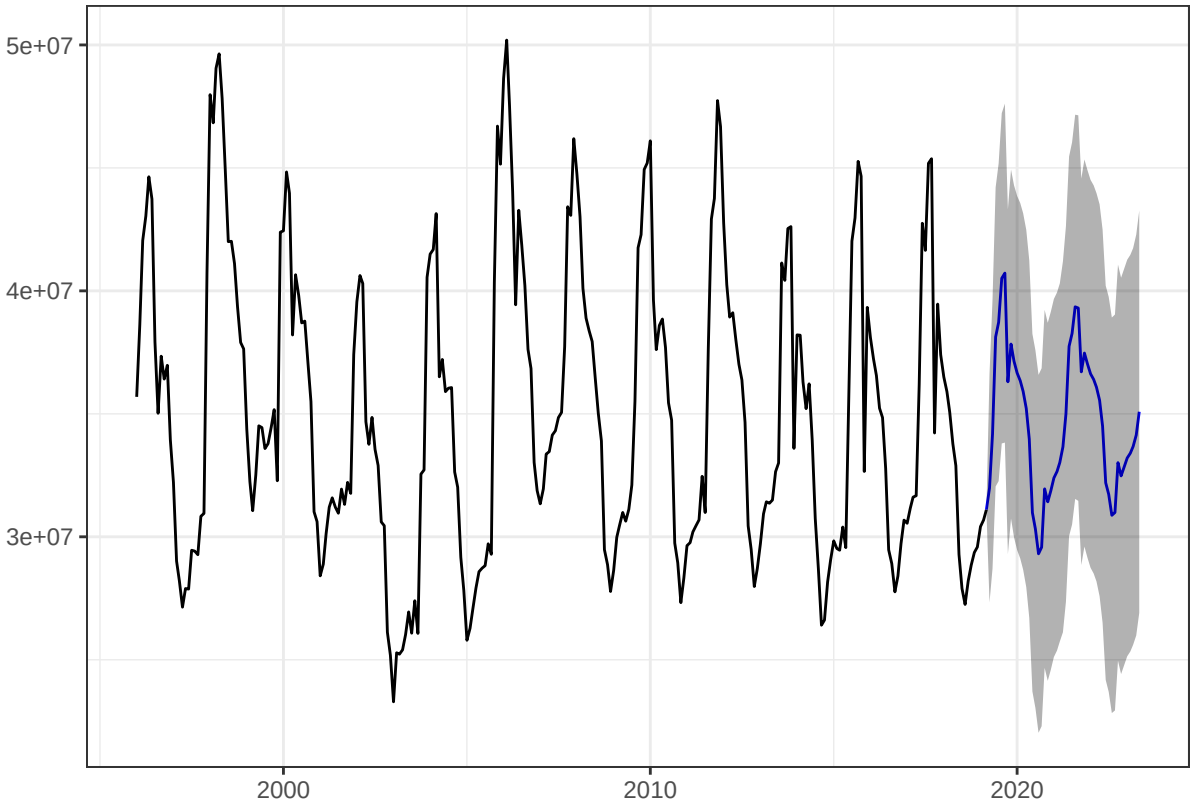

Figure 5: Forecasting