Data Analysis Skill Test
4Intelligence

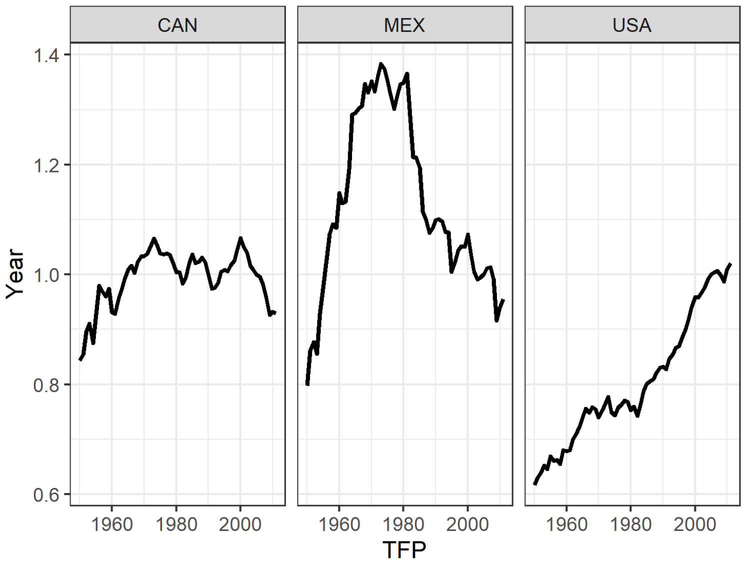
By Candidate JÚLIA GHIZZI

December 2020

I. Case 1

Part I From Figure 1 one can guess that the three time series are not stationary. Also, plotting x_t against $x_t + 1$ [1] generates a could of observations forming straight line with positive slope for all time series, and such positive correlation remains strong for the USA TPF series as far as lag 10.

FIGURE 1: TFP TIME SERIES: MEX,CAN AND USA



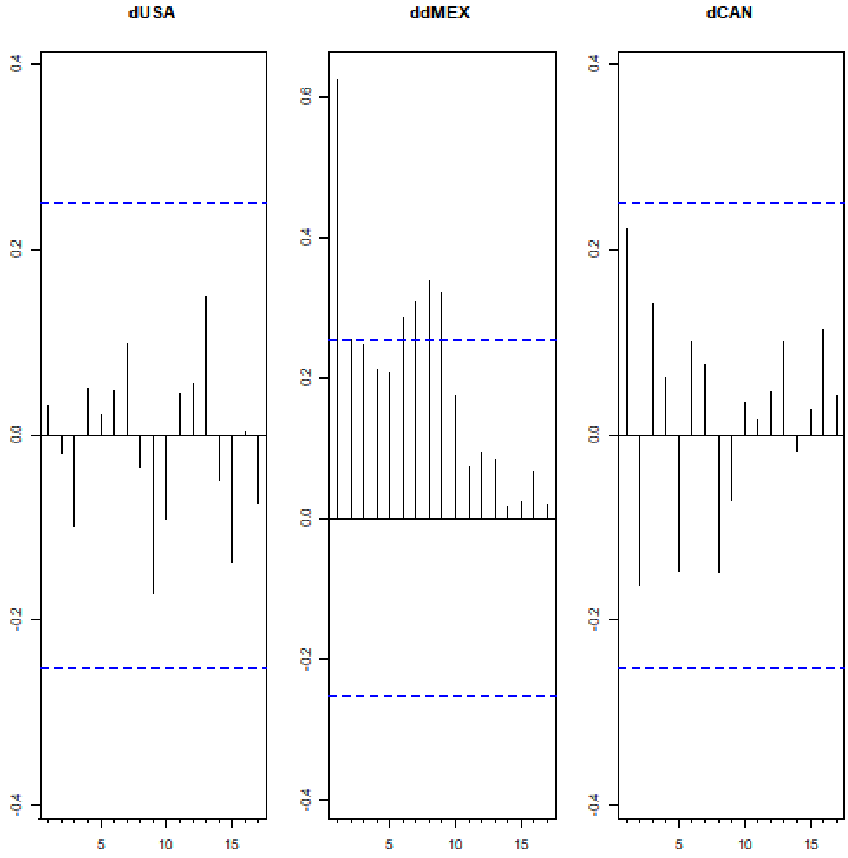
In the three cases, one fails to reject the null hypothesis of a unit root in the augmented Dickey-Fuller test and in the Phillips-Perron test. Taking the first difference is enough to handle the stochastic trend of the USA and the CAN time series, while the MEX time series does not have a unit root after taking the second difference.

Since the ACF and the PACF in Figure 2 of the first difference of the USA TFP indicates that the TFP growth is a random walk, the original series might be a random walk with a drift. The process underlying the Canadian and the Mexican is not that clearly verified by graphical analysis. The initial guess would be that MEX TFP is generated by an ARIMA(p,2,0) once its ACF declines slowly and the PACF is truncated., and CAN TFP is an ARIMA(0,1,q).

[1] Plots are not included in the report but R script generates them.

FIGURE 2: CORRELOGRAMS FOR THE STATIONARY SERIES

(A) Autocorrelation Function



(B) Partial Autocorrelation Function

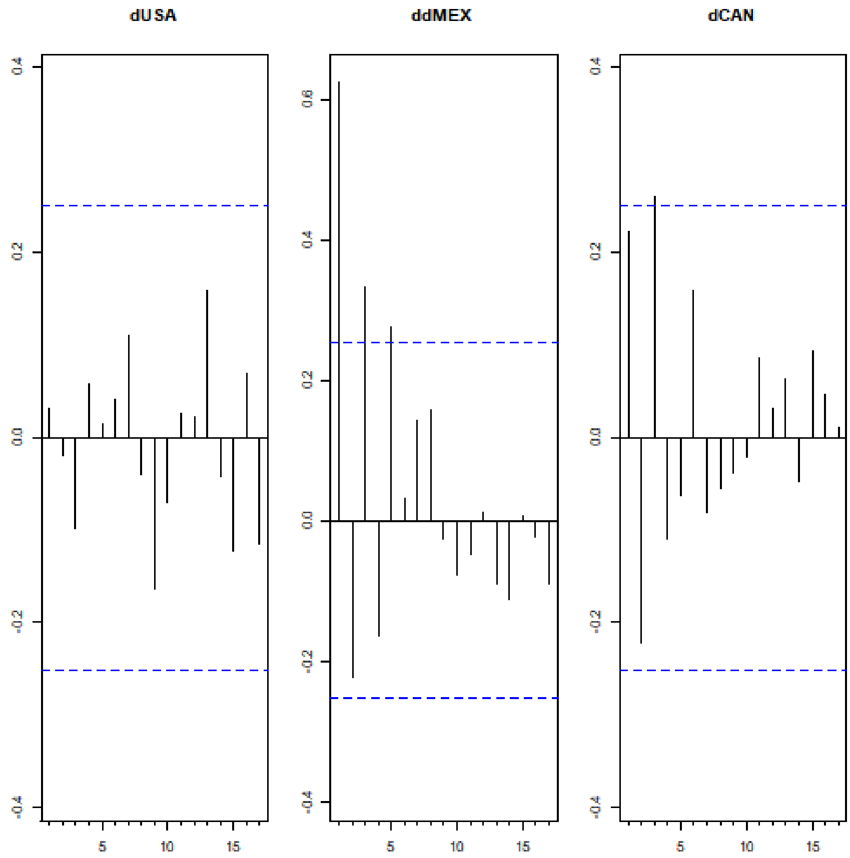
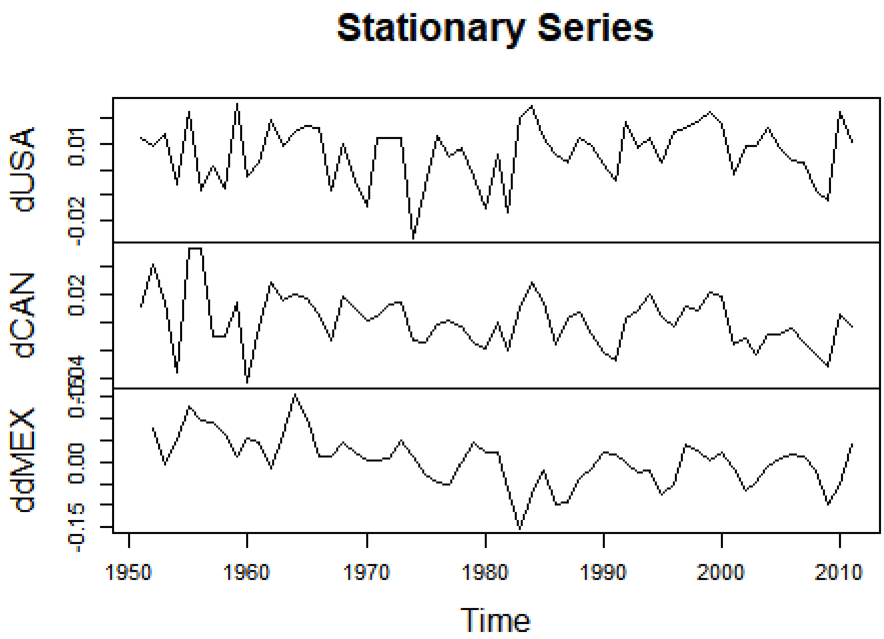


FIGURE 3: STATIONARY DIFFERENCED TFP TIME SERIES: MEX,CAN AND USA



Part II Since there are no covariates available, I follow the Box-Jenkins methodology to forecast the TFP series. The initial analysis of each time series provided material to formulate initial guesses.

The training sample is from 1960 to 2001, the last 10 years are used to verify the quality of the predictions. Using the R function *auto.arima*, that selects the ARIMA(p,d,q) model that better fits the provided training time series according to information criteria, I select an initial model. The parameter d is set to the number of differences necessary to achieve stationarity. Other parameters are free.

As a result, the TFP series of USA is initially modeled as an ARIMA(0,1,0) with drift, the TFP series of MEX is modeled as an ARIMA(0,2,1) and the TFP series of CAN is modeled as an ARIMA(2,1,2).

For each initial model, the following were verified: (i) coefficients' significance, (ii) residuals normality, (iii) residuals stationarity based on graphical analysis, and (iv) the fit of fitted value against the values of the training sample. All models perform well in terms of fitting past data.

However, the forecasts for 2002 to 2011 aren't good for the Candian and the Mexican time series. Hence, I have searched around the initial model. I search for the ARIMA model with AR parameter p in $\{p*-2, p*-1, p*, p*+1, p*+2\}$ where $p*$ is the AR parameter of the ARIMA model selected by *auto.arima*. Similarly, I also allowed the MA parameter to vary within 2 units above and beyond $q*$. The criteria to define the 'best model' around the initial model are the RMSE and the MAE for the forecasts for 2002-2011.

Ten year forecasts of the initial model and the 'best' model according to RMSE and MAE criteria are displayed in Figure 4 and Table 1. The confidence interval for the forecasts of the CAN and the MEX time series is quite large. Nonetheless, if the long-term forecasts appear erratic or unstable, there may still be some stochastic trend, which is not the case.

TABLE 1: 10 YEAR FORECASTS OF USA, CAN AND MEX TFP

(A) Initial Model

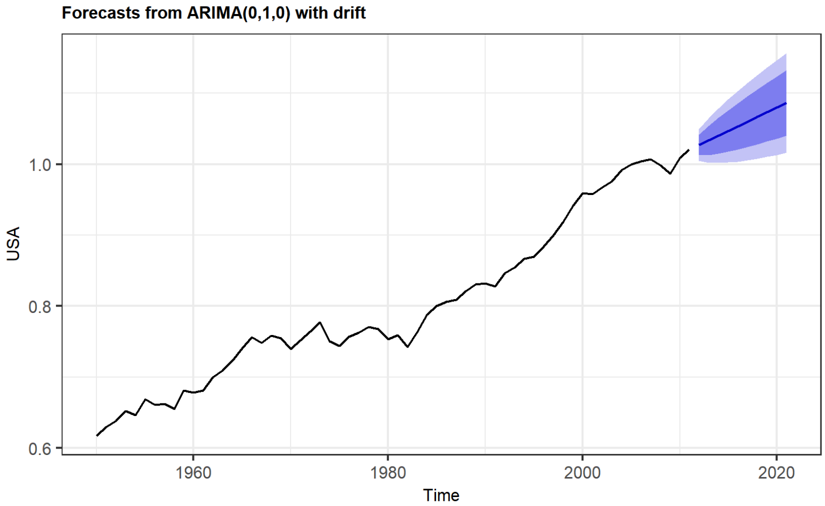
Year	USA	CAN	MEX
2012	1.027	0.920	0.949
2013	1.034	0.920	0.943
2014	1.040	0.926	0.937
2015	1.047	0.921	0.930
2016	1.053	0.921	0.924
2017	1.060	0.924	0.918
2018	1.067	0.922	0.911
2019	1.073	0.922	0.905
2020	1.080	0.923	0.899
2021	1.086	0.922	0.892

(B) 'Best' Model

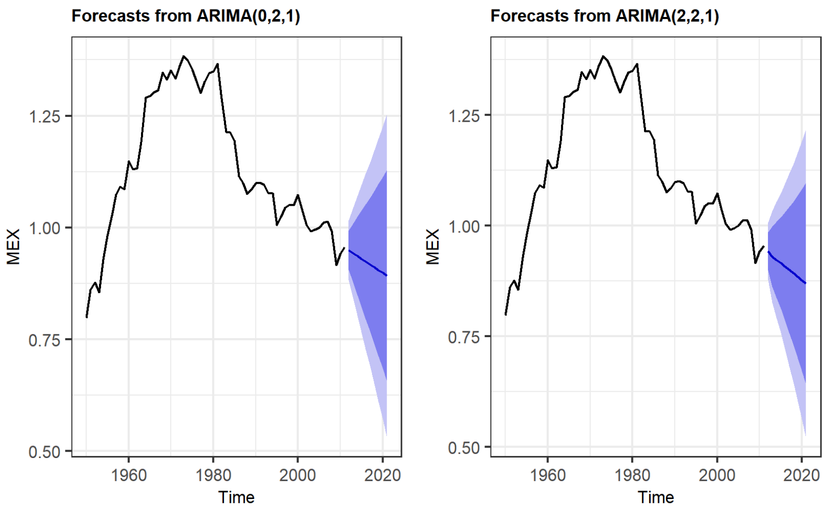
Year	CAN	MEX
2012	0.921	0.942
2013	0.922	0.930
2014	0.923	0.923
2015	0.918	0.916
2016	0.916	0.908
2017	0.919	0.900
2018	0.915	0.892
2019	0.913	0.885
2020	0.915	0.877
2021	0.913	0.869

FIGURE 4: 10 YEAR FORECASTS OF USA, CAN AND MEX TFP

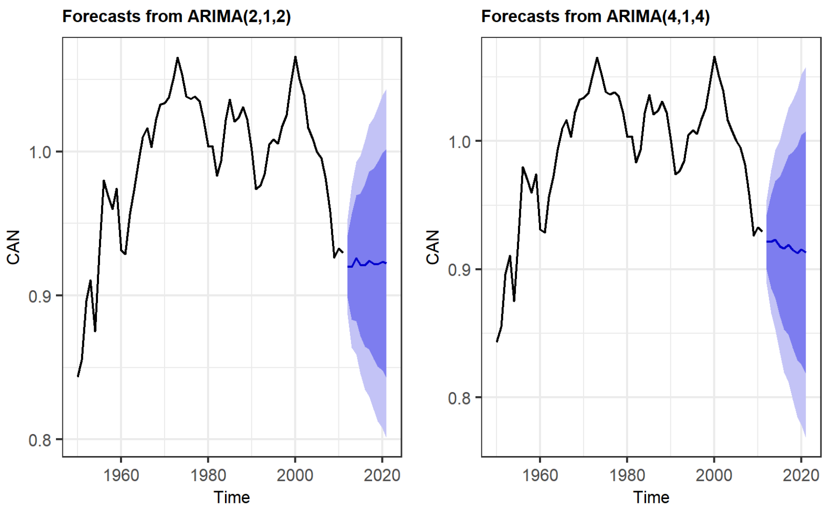
(A) USA



(B) MEX



(C) CAN



Part III The TFP is a measure of productivity build upon data on production inputs such as capital stock, human capital, and the number of workers in the economy. Intuitively, it is the part of the growth in value-added that can not be explained by the changes in inputs. Hence, for forecasting purposes, having data on capital stock, human capital, labor force and output would allow us to track observed changes in inputs and outputs disentangling those changes from TFP changes. Therefore, I consider adding the following variables

- Real GDP
- Number of persons engaged (in millions).
- Average annual hours worked by persons engaged.
- Index of human capital per person, based on years of schooling (Barro and Lee 2013) and returns to education (Psacharopoulos 1994)
- Capital stock

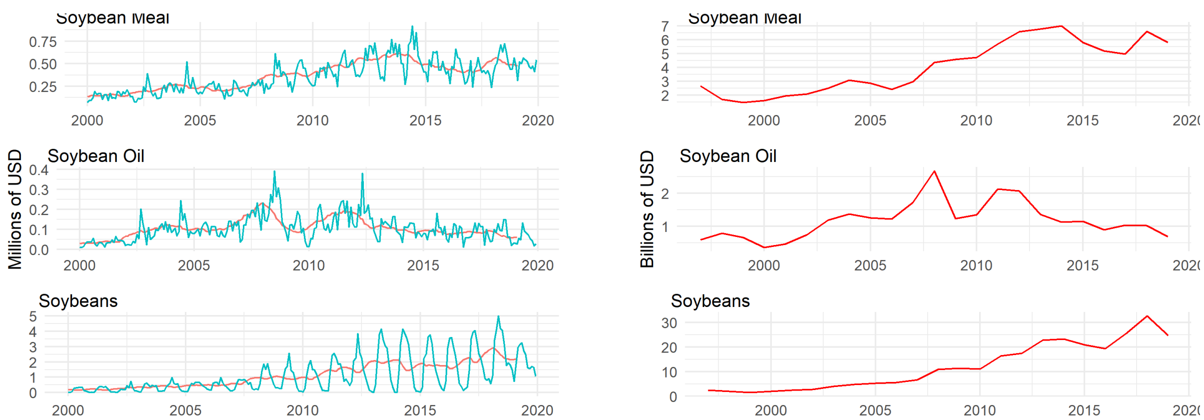
The specific modeling technique would depend on the characteristics of these variables.

II. Case 2

Part I The following figures illustrate the evolution of total value Brazilian exports of soybeans, soybean meal, and soybean oil. The monthly time series displays a strong seasonality, thus, a 12-month moving average [2] is included in panel A to get a clear view of exports time trends. Given the simple smoothing technique, the monthly moving average and the pattern of the annual time series are very similar.

FIGURE 5: EVOLUTION OF BRAZILIAN EXPORTS OF SOYBEANS AND ITS PRODUCTS

(A) Monthly Data (B) Annual Data



[2] To preserve the end observations of the time series, the rolling average of some month t includes observations of the 11 previous months $\{t-1, t-2, ..., t-11\}$

Part II It is evident from Figure 6 that soybeans exports steadily accounts for the majority of the export value (56.77% in Table 2) and volume in the last five years. Soybeans exports are followed by corn (16.35%) and soybean meal (13.30%) in the ranking of the most important products. From 2014 to 2019, corn exports rose its share in total export value in 7.59 p.p. and the volume exported more than doubled in that period. Meanwhile, soybean meals exports followed the opposite direction, with its share in total export value diminishing in 2.32p.p. in the same period. Nonetheless, is not likely that soybean meal will fall in the ranking once the fourth place is not in an upward trend. Sugar exports, the fourth place in the export value ranking, have decreased its share in total export value by 9.21 p.p. from 2014 to 2019, and part of this shift is due to a 26.3% reduction in the amount of sugar sold to the rest of the world. Even though soybean meal exports lost some of their relative importance, the volume trade is still increasing. Therefore, the most important products of Brazilian exports are soybeans, corn and soybean meal.

FIGURE 6: COMPOSITION OF EXPORTS

(A) Volume (B) Value

]

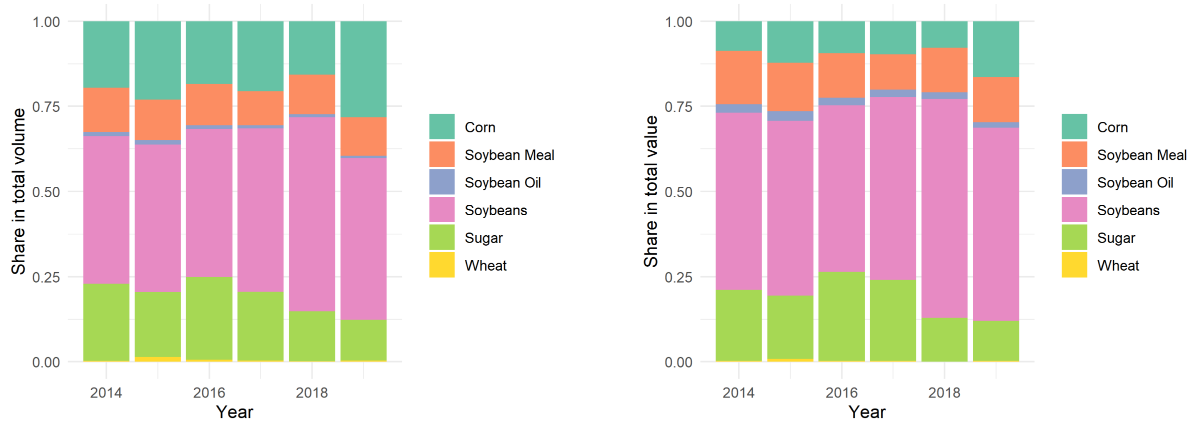


TABLE 2: PATTERNS OF EXPORTS GROWTH BETWEEN 2014 AND 2019 BY PRODUCT

Product	Gr. of Volume	Gr. of Value	Change in Share	Share in 2019
Corn	102.77%	81.49%	7.59p.p.	16.35%
Soybean Meal	20.52%	-17.2%	-2.32p.p.	13.3%
Soybean Oil	-20.89%	-38.99%	-0.94p.p.	1.58%
Soybeans	53.87%	6.29%	4.84p.p.	56.77%
Sugar	-26.3%	-45.52%	-9.21p.p.	11.74%
Wheat	100.24%	14.03%	0.04p.p.	0.26%

Part III Panel A of Figure 7 reports the share of corn exports that were transported through the air, ground, river, sea, and other means of transportation. For comparisons across years (Panel A), the figure displays patterns of exports volumes across years and for comparison across origins and destination in a given year (Panels B to D) figures illustrate the patterns of the value of exports.

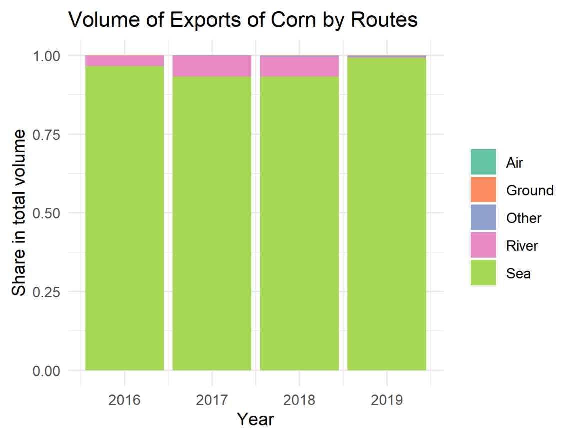
Clearly, in the last five years, corn exports were mainly delivered by sea. This pattern is actually consistent across the commodity products in the database. The relative importance of the routes is heterogeneous in the Brazilian States. In Acre, the main route to export corn is through the ground, and in Pernambuco, the only route is through the air. Acre and Pernambuco exports account for a very small share of total corn exports.

Also, the main route of corn, and other products as well, might shift according to the destination. When exporting to top patterns (Panel A of Figure 7), the ones the purchased the greatest amounts of corn, the sea remains the main route. However, when exporting to the countries that bought the smallest amount of corn (Panel B of Figure 7), one of the corn's main routes in through the air.

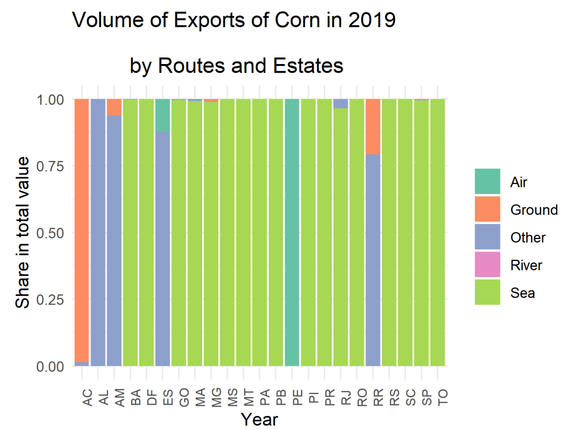
Overall, the main route of corn is the sea. There is some variation across State which could be driven by local infrastructure, and some variation across buyers that could be explained by the cost of transportation. Larger amounts are exported through the cheapest route, while small amounts are delivered by an expensive route, the air.

FIGURE 7: ROUTES OF CORN EXPORTS

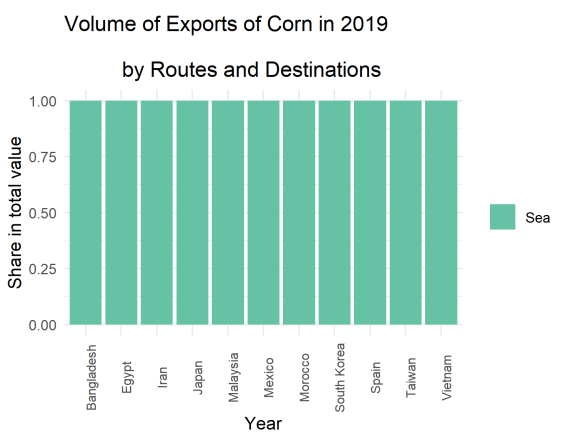
(A)



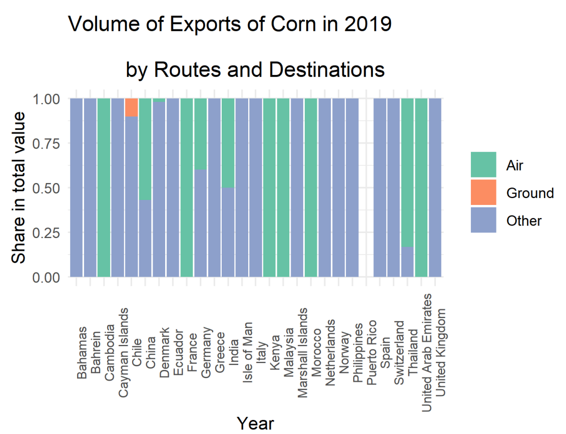
(B)



(C) Top Buyers of Brazilian Corn



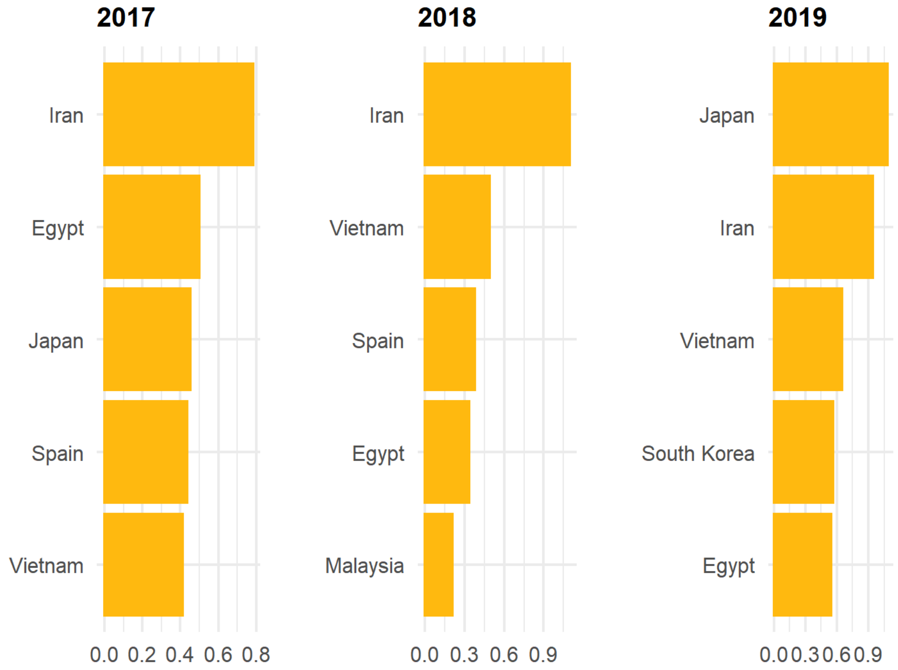
(D) Bottom Buyers of Brazilian Corn



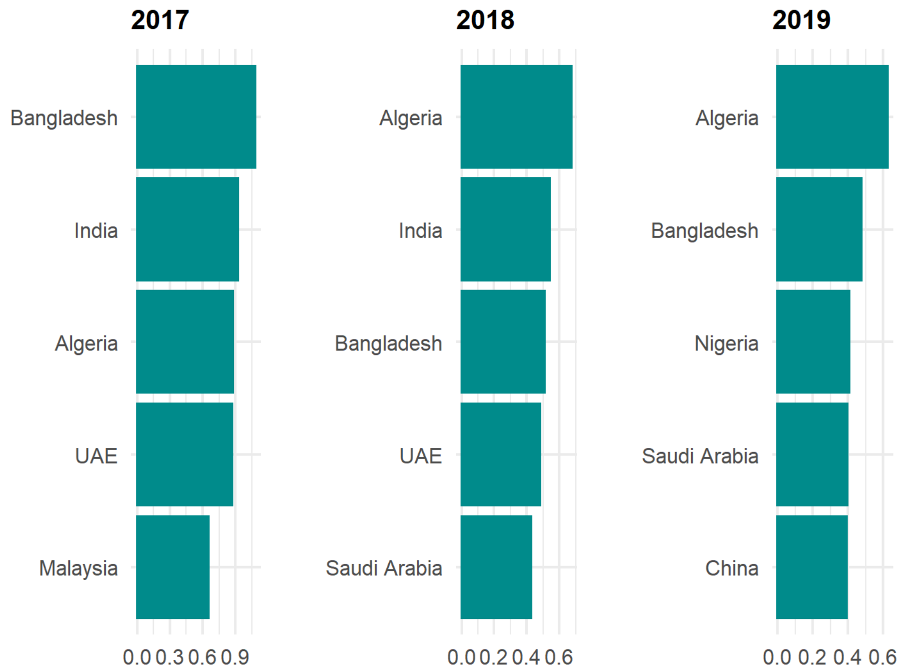
Part IV Figure 8 reports the most important trade partners in the markets of corn and sugar. The criteria to build the ranking is the volume traded each year.

FIGURE 8: MOST IMPORTANT TRADE PARTNERS IN THE CORN AND IN THE SUGAR MARKETS

(A) Top Importers of Corn



(B) Top Importers of Sugar



Part V Figure 9 reports the most important States in terms of exports. The criteria to build the ranking is the average volume traded in the last three years, that is from 2016 to 2019.

FIGURE 9: TOP EXPORTERS