Data Analysis in the Airline Industry: Revealing the Impacts of COVID-19 on Global Air Transport (and Vice Versa)

Paula Giovanna Rodrigues[1], Victor Antonio Menuzzo[1], Semida Silveira[2], Ricardo Lüders[1], Luiz Gomes-Jr[1]

[1]Departamento de Informática – Universidade Tecnológica Federal do Paraná (UTFPR)
Curitiba – PR – Brazil

[2]Systems Engineering Program, Cornell University Ithaca, NY 14853 – USA

paularodrigues@alunos.utfpr.edu.br

Abstract. *The aviation sector is a major factor in our globalized society, funding the highest levels of advanced engineering for its own operations and providing irreplaceable services to various industries such as tourism and transportation. On the other hand, the efficiency and scale in which people are moved through air traffic plays a central role in the spread of diseases such as COVID-19. This paper analyzes the interactions between air traffic and the current COVID-19 pandemic. By combining air traffic with epidemiological data we assess (i) how COVID-19 affects the aviation industry, including factors that explain the differences in impact among countries, (ii) how the transportation of passengers contributes to the spread of the disease globally, and (iii) how air traffic restrictions are related to epidemic status in the countries. We explore these issues with a combination of exploratory analysis, modelling and complex network analysis. We believe that our models and analyses contribute for developing guidelines to minimize impacts in health and economic outcomes.*

1. Introduction

The airline industry is an important player in the transportation sector, promoting efficient and safe means for persons and goods to move across the globe. The Industry High Level Group (IHLG) reported recently [IHLG 2019] that airlines carried around 4.3 billion passengers annually and 58 million tonnes of freight. The volume and speed in which people are transported by air also make airlines a central player in the spread of infectious diseases such as COVID-19. The effect of air transport in the early dissemination of COVID-19 has been assessed in previous works [Lau et al. 2020, Chinazzi et al. 2020]. Being an enabler of long distance transportation and connecting the whole world, the airline industry is itself vulnerable to pandemic events that require people to minimize travel. The sector has seen major impacts from previous pandemics, but losses due to the current crisis can be substantially higher [Iacus et al. 2020]. A better understanding of the interplay between air transport and the COVID-19 pandemic is important to minimize health and economic costs, especially considering that developing countries are expected to have greater losses [World Bank 2020].

This paper explores these relationships, analyzing factors that influence the magnitude of impacts in different countries. Our main research questions are organized in both directions of the relationships being analyzed as follows: i) COVID-19 effects on air transport: What factors influence the magnitude of air traffic losses in different countries?; ii) Air transport effects on COVID-19: How the influence of air traffic on the dynamics of the pandemic changes over time? iii) Combined effects on countries: Are airline losses directly related to epidemic risks?

To better understand the relationship between airline industry and the current pandemic scenario, we employed several data science techniques, including exploratory analysis, complex networks, model optimization, and regression analysis. We show that there is unbalanced impacts of the pandemic on different countries. Differently from other works of the literature, we focus on a bidirectional relationship between COVID-19 effects and air transport reduction. Our main contribution is showing that there is no direct relationship between epidemic concerns and air traffic losses. This result is based on a new risk metric, named virality potential. These findings suggests that data-driven responses to pandemic events could be more effective for both preventing spread of the pathogen and minimizing economic impacts.

This paper is organized as follows. Sections 2 and Section 3 present the related work, and data sources, respectively. Section 4 analyzes the air traffic data, and Section 5 focuses on epidemic dynamics influenced by air traffic. Section 6 presents the regression analysis to assess the influence of virality on losses.

2. Related Work

Recognizing the importance of the aviation industry and its role in the spread infectious diseases, the relationships between air traffic and COVID-19 have been studied since the early days of the pandemic. The studies often focus on one aspect of the bidirectional relationship, either considering the economic impact of air traffic reduction due to COVID-19 or evaluating the containment of COVID-19 epidemic spread through air routes. Iacus et al. [Iacus et al. 2020] use historical flight and online booking data to build a forecast model used to estimate the number of lost passengers and its economic impact in different scenarios. The authors estimate losses in a short recovery scenario (12 months to restore pre-pandemic levels) and a worst case scenario where the industry does not fully recover (reaching only 60%). The authors predict global impacts ranging from 12.81 to 30.31 million jobs and 0.70 to 1.67 percent points of GDP loss. With the dramatic and quick loss in air traffic caused by the COVID-19 crisis, it is expected that the sector will have a long path to recovery. Sobieralski [Sobieralski 2020] estimates a reduction between 7% to 13% of the airlines' workforce in a post-Covid-19 scenario. The most affected positions are expected to be in occupations dedicated to passenger handling.

The impact of air traffic on spread of Sars-Cov-2 virus has also been measured using different strategies. Lau at al. [Lau et al. 2020] analyzed data on air traffic and passengers transported in Chinese and international routes. The authors found high correlations between transported passengers and contamination by COVID-19 inside and outside regions of China. To reduce the rate of contamination, governments have implemented air travel restrictions. Chinazzi et al. [Chinazzi et al. 2020] measure the effectiveness of the policies using a global epidemic and mobility model (GLEAM). The model simulates the movements of people taking into account air and ground routes. The authors highlight the significant impact of travel ban in Wuhan-China as an initial 77% reduction in cases imported by other countries. Their findings suggest that, although air traffic plays a major role in the initial spread of the pandemic, other measures of containment (e.g. social isolation) are more effective in long term than air traffic restriction.

A country's health capacity strongly influences the management and control of COVID-19 case imports. Gilbert et al. [Gilbert et al. 2020] used data on the volume of air travel departing from China and directed to Africa to estimate the risk of importation per country. Their study used two indicators to determine countries' capacity to detect and respond to cases: preparedness, to measure how well a country is prepared to deal with the virus; and vulnerability, to measure how likely a country is to be infected with the virus. Their analysis found that countries with highest import risk have more capacity to respond to outbreaks than countries with moderate risk. A global action to push social distancing has gained strength with COVID-19 pandemic, which thus affects the transportation sector. Nakamura and Managi [Nakamura and Managi 2020] simulated three different scenarios for the risk of importing and exporting the virus based on world airports by changing the spread rate (between 0.5 and 1.0) and air traffic reduction of 90% for countries in the 1st quartile of incidence area, 60% in the 2nd, and 30% in the 3rd quartile. The authors concluded that the risk is still high even with restrictions, indicating that a strong reduction is necessary in air traffic (i.e. higher than 90% for countries of the first quartile). Moreover, public health measures are necessary in international travels that prevent reintroduction of COVID-19, especially when internal country restrictions are relaxed. Dickens et al. [Dickens et al. 2020] created six different structures to calculate the number of infectious and future infectious travelers. Their results show that there will be a spread reduction of 91.2% by screening all travelers and isolating cases for at least 14 days in relation to non-screening.

The aforementioned works show the importance of aviation industry in the spread of COVID-19 and the respective economic impacts. However, most of them shed light on separate aspects of the spread disease and economic impacts on aviation industry. In this paper, we aim at moving forward by contextualizing the impacts in a globalized world, and emphasizing the differences in health risks and economic impacts among countries.

3. Datasets and methods

In this work we use data on epidemic outcomes, flight statistics and socioeconomic indicators provided by multiple sources as follows. Airports connections are obtained from OpenFlights (https://openflights.org), which also provided data on type of flights (international or domestic) and conversion tables for airport codes to country names. Information on flight date, airport of origin and destination (historical flight tracking) were sourced from OpenSky (https://opensky-network.org). To perform our analyses, we used data from 01-01-2020 to 05-31-2020, corresponding to a total of 108,786,427 flights. To calculate the percentage of loss in flight volume, we used averages of January, 2020 as baseline. We chose January as a baseline since previous months had irregular data in the source; and since we are computing percent loss, we believe that seasonal fluctuations will not have a significant impact in the results. Both origin and destination were used for the volume of flights by country. Flight data with missing information on origin or destination were discarded. To calculate the percent loss of flights, we used national and international flight data, but in the contagion model of Section 5, only international flights were used. Data from World Bank Group (https://www.worldbank.org/) provides socioeconomic indicators for countries, namely gross domestic product (GDP), per capita gross domestic product, export of goods and services, and population size. Data on the number of COVID-19 cases (COVID-19 reports) for each country was obtained from Our World in Data (https://ourworldindata.org).

To assess the losses in air traffic and their impact on countries, we start with an exploratory analysis (Section 4). To capture the non-linear epidemic behavior we employed a stochastic model of the diffusion of the virus through the airline network (Section 5). Motivated by the observation that epidemic risks were not the main factor in flight losses, we propose a new metric to measure the likelihood of countries exporting infected cases through the airline network, and employ regression analysis to contextualize some of the factors discussed throughout the paper (Section 6).

4. Impact of the pandemic on airline industry

This section focuses on measuring the losses absorbed by the airline industry as a consequence of the pandemic. Calculating overall economic losses is a complex task, since there are many variables involved. Reduction in flight ticket demand, due to formal travel bans or individual choice, is a major factor in the crisis. However, countries and companies can cope differently with the challenges, since in some situations cargo transportation can be increased (for example, by shipping much needed medical supplies). Also, different countries have different means to deal with companies in economic distress and to regulate layoffs of employees. Lower prices of fuel have also reduced some losses. Here we do not aim at including all variables to estimate economic losses. Instead, we focus only on flight volume and how it changes in time for different countries. We expect that the relative flight volume regarding to the normal operation represents a reasonable proxy for the overall percentage losses.

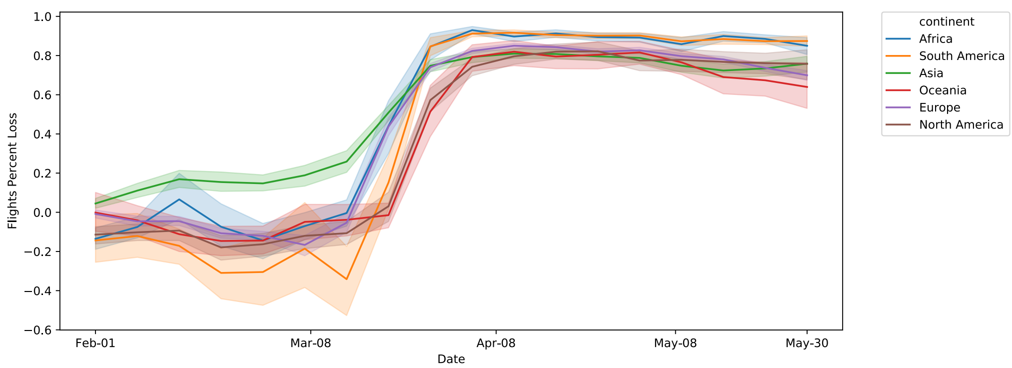

Figure 1. Average variation of flight volume by continent (02/01 to 05/30/2020)

We start with an exploratory analysis of losses along time. Figure 1 shows the average reduction in flight volume by continent. According to Figure 1, Asia was the first continent to experience losses, but it was soon followed by other regions. Figure 2 shows volume reduction for selected countries (based on their viral potential – detailed in Section 6 – and their role in the pandemic). China experienced significant losses at the beginning of the pandemic, but started its recovery after a few weeks. The United States has had smaller losses compared to other countries. The reduction in traffic detected in

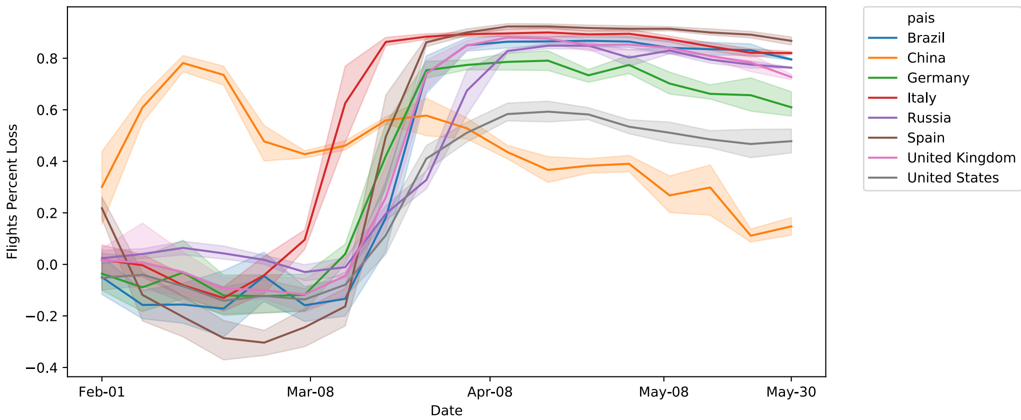

Figure 2. Average variation of flight volume by country (02/01 to 05/30/2020)

our analysis is consistent with the reports from the Official Aviation Guide (OAG)[1] with an overall 0.75 correlation. The reduction in flight volume is obviously caused by the pandemic, but our hypothesis is that epidemic dynamics cannot explain the variance in the intensity of loss among countries.

We analyzed several variables to better understand the variation in flight reduction, including Gross Domestic Product (GDP), GDP per capta, population, exports of goods and services. The strongest correlations with air traffic losses were identified for exports of goods and services (correlation of -0.52 with p-value < 0.01), and GDP (correlation of -0.51 with p-value < 0.01). These results suggest that stronger economies experienced smaller losses, which can be due to cargo transportation or particular passenger flights (e.g. business or repatriation).

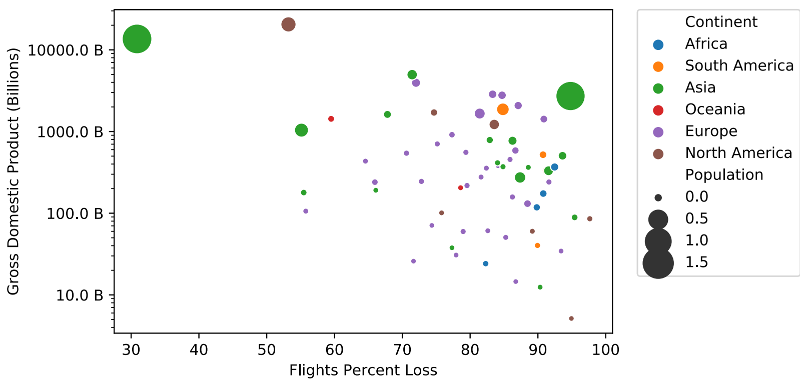

Figure 3. Flight volume loss $\times$ GDP

Figures 3 and 4 show countries according to GDP, and exports of goods and services, respectively. Each point of the plots corresponds to a country in the respective continent whose bullet is proportional to the country's population. By comparing these

[1]https://www.oag.com/coronavirus-airline-schedules-data

two variables (GDP, and exports of goods and services) with flight percent losses, there is a trend of poorer countries experiencing higher losses.

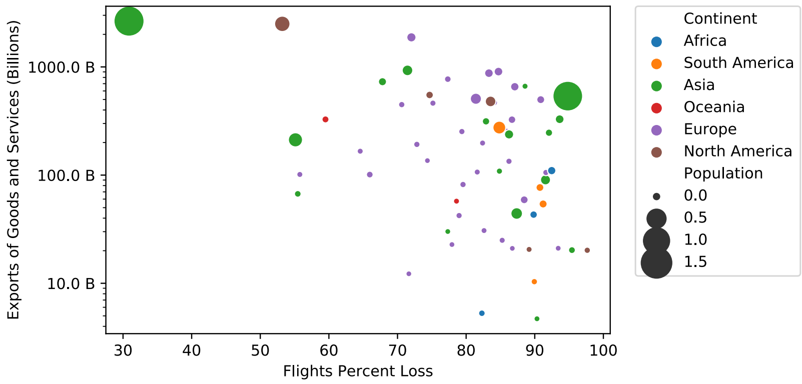

Figure 4. Flight volume loss × exports of goods and services

The analysis of this section show the direct influence of the pandemic on airline industry with unbalanced impacts on countries. The opposite is considered in the next section, by assessing how air traffic impacts the pandemic spread.

5. Air transport and epidemic dynamics

Airline routes have a major role in the spread of an epidemic across regions as expected and verified in [Lau et al. 2020, Chinazzi et al. 2020]. The underlying factors are the efficiency of airlines for transporting passengers across different regions and how countries are increasingly connected through airline routes. Figure 5 shows the connectivity among countries based on our data sources. The width of an edge connecting two countries is proportional to the average number of flights per week between them in January, 2020.

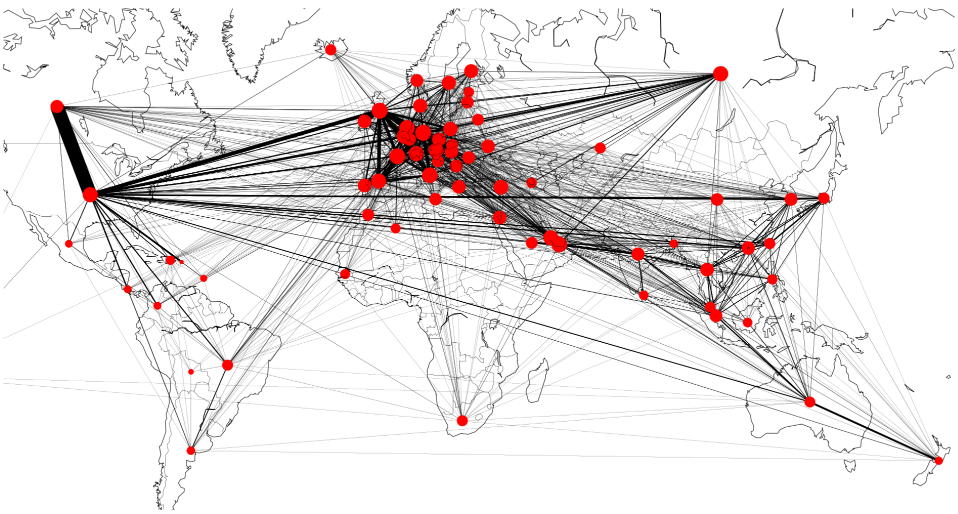

Figure 5. Air routes network used in the analyses.

We built two epidemic models for assessing the evolution of air traffic influence on spreading COVID-19. The baseline model considers the levels of air traffic just before the pandemic (January 2020) while the other model takes into account the air traffic volume losses between countries. Both models are fitted to epidemiological data (daily reported COVID-19 cases) in order to assess how effective they are for explaining the pandemic evolution. The models only capture the influence of airline routes and air traffic volume reductions. The assessment of other variables such as containment policies (social isolation, quarantines, mask usage, etc.) and population dynamics (density, mobility, etc.) is out of the scope of this paper (these aspects were covered in, e.g. [Zhang et al. 2020]).

We specified a model to capture the relative importance of air traffic in the pandemic. In the general model, each state represents one day since the beginning of the pandemic (the first official cases were registered on Dec 31, 2019). A model for diffusion of cases is given by Eq. (1). Each component of vector $\mathbf{u}_t$ contains the number of cases in a country on day t, which grows at rate r_t. The cases have a scalar probability p_t of traveling to another country by following the weighted transition matrix M_t which represents the number of flights of connecting routes between countries updated on day t.

$$\mathbf{u}_t = r_t \cdot \mathbf{u}_{t-1} + p_t \cdot \mathbf{u}_{t-1} M_t \tag{1}$$

The baseline model presented above is branched out into two models *NoLoss* and *WithLoss*, which update M_t with the number of flights with and without reduction due to the pandemic, respectively. The computed reduction is shifted by two weeks to capture the COVID-19 incubation and viral load peak as well as the delay for testing and reporting cases. One pair (r_t, p_t) is computed for each day t after February 1st by using least-square minimization of f_t given by Eq. (2).

$$f_t = \arg\min_{r_t, p_t} \|\mathbf{c}_t - \mathbf{u}_t\|^2 \tag{2}$$

The goal is to estimate how well each model fits real data as time progresses. Each component of vector $\mathbf{c}_t$ contains registered cases of COVID-19 in a country for day t. The obtained models have $r \in [1.1, 1.2]$ and $p \in [10^{-2}, 10^{-4}]$. We calculate the *Spearman* correlation between vectors of estimate and real cases of COVID-19 for each day t to analyze how well the models reflect the observed data. We do not use *Pearson* correlation because it is very sensitive to outliers – the number of cases in China at the beginning of the epidemic is much larger than all others, which results in very high ($corr > 0.98$, $p < 0.01$) but unrealistic *Pearson* correlations.

Figure 6 shows the evolution of correlations for each model. The prediction of the infection rankings of affected countries are good for both models at the beginning of the pandemic due to high correlation values. The correlation drops as time passes, which indicates that imported cases have smaller impact after local contamination starts as expected. However, one interesting aspect is that correlations provided by the model which includes air traffic losses drop even more than those of the baseline. It indicates that the observed rates of air traffic are not optimal in containing the pandemic.

6. Regression analysis of air traffic loss $\times$ country virality and GDP

Considering the clear effects identified in the previous sections, we now assess how epidemic risks and air traffic losses are reflected in countries. According to the analysis of

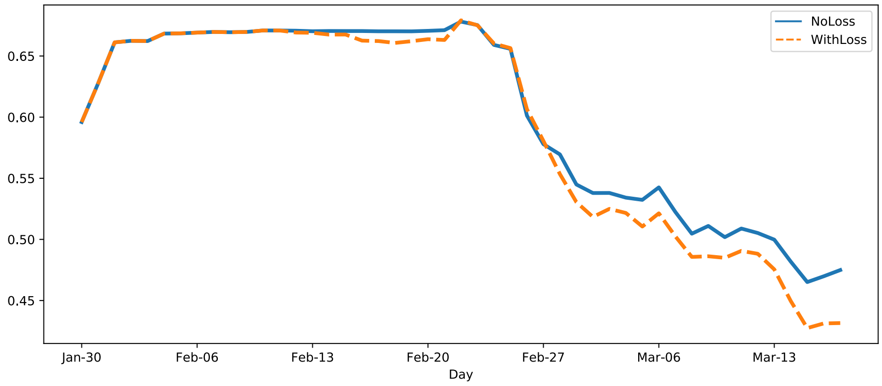

Figure 6. Spearman correlation between vectors of estimate and real cases of COVID-19 for models *NoLoss* and *WithLoss*.

Section 5, it is noticeable how the effects of international travels on the pandemic tend to become less pronounced over time. Ideally, risk reduction should guide the recovery of airline industry based on an effective risk assessment.

In this section, we further explore the dissociation between long term evolution of the pandemic and the observed reduction in air traffic at country level. To assess risk, we developed a compound metric to measure the potential of a country spreading the virus to other countries. The metric combines (i) the status of the epidemic (how widespread is the virus) in each country; (ii) the country's importance in the air-traffic network (is it an important hub for passengers?). The metric is given by Eq. (3).

$$v_t(c) = \frac{E(c)A_t(c)}{\max_{c \in C}\{v(c)\}} \tag{3}$$

For all countries $c \in C$, our metric $v_t(c)$ assesses the virality score of c on day t. The eigenvector centrality $E(c)$ represents how well connected c is in terms of flight routes, and $A_t(c)$ accounts for the number of active cases of c in a given day t. Figure 7 shows the virality potential of countries along time for the countries with highest virality.

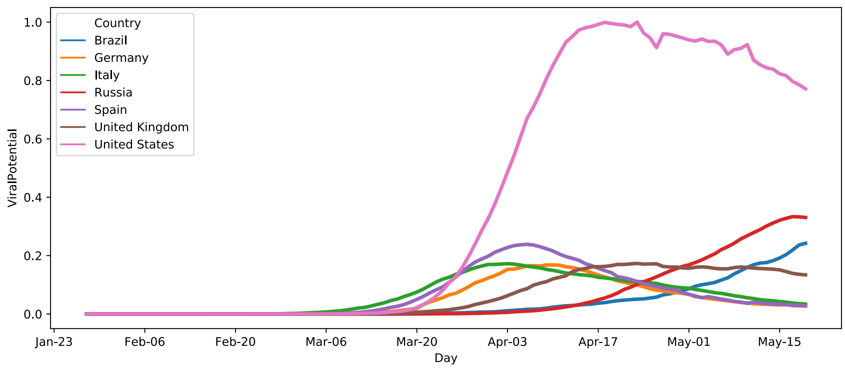

Figure 7. Virality potential of countries along time

We use $E(c)$ as a measure of connectivity for c according to the graph of Figure 5. The eigenvector centrality [da F. Costa et al. 2007] is a metric used in complex network analysis that assesses connectivity recursively (basically, a connection with a well connected node transfers more connectivity value to the original node). The number $A_t(c)$ of active cases of COVID-19 in a country is estimated by aggregating new reported cases in the 14 days leading to t. The metric for *viral potential* is normalized by the maximum value (for all values to be between 0 and 1).

We performed a series of regression analysis to further quantify and contextualize the patterns identified in the previous sections. Models 1 and 2 of Table 1 refer to Eq. (4). The time index refers to weeks between 2020-04-11 and 2020-05-16.

$$loss_t = \beta_0 + \beta_1 \cdot GDP + \beta_2 \cdot Virality_t \tag{4}$$

Model 2 includes entity fixed effects, capturing factors that associated with specific countries that are constant over time. The inclusion of entity fixed effects significantly improves the variance captured by the model. This suggests that factors such as policies imposing flight restrictions and voluntary avoidance of flights are the most important aspects. These analyses show once more the disconnection between the behavior of airline industry losses and epidemic factors.

Table 1. Coefficients of regression models 1 and 2.

Covariate	Model 1	Model 2
GDP	$-3.4E{-}12^{*}_{2.9E-13}$	$-3.1E{-}12^{*}_{4.6E-13}$
Virality Potential	$46.20^{*}_{7.55}$	$16.81_{13.54}$
Adjusted R^2	0.296	0.803
Fixed Effects	-	Country

Subscripts are standard errors; * significance level of 1%.

7. Conclusion

In this paper we explored the interactions between airline industry and COVID-19 pandemic. It was carried out with a combination of exploratory analysis, diffusion modeling and complex network analysis to understand the associations between flight traffic losses and the Covid-19 pandemic. Our analysis suggest that there is a dissociation between air traffic and epidemic dynamics as time progresses. Revisiting our research questions, we observe that: i) The magnitude of air traffic losses in different countries is associated with their GDP and volume of exports, with developing countries experiencing higher losses in the airline industry; ii) The influence of air traffic on the dynamics of the pandemic changes over time, with the effect becoming less relevant as time progresses; iii) Air traffic losses are not directly related to epidemic risks, as evidenced by countries with highest viralities not experiencing consequences proportional to their epidemic potential. The analysis has limitations as well. Our dataset for flights seems to have irregular data for small countries. Therefore, we excluded 48 of these countries which represent 0.5% of total number of flights. Furthermore, we could not differentiate cargo from passenger flights, which will distort the results especially for countries that export valuable goods. The correlations between air traffic loss and economic variables require further statistical analysis to support meaningful conclusions. We are using reported cases to estimate cases of COVID-19 which do not reflect the total of cases because only part of the population is tested and testing rates vary widely among countries. Better estimates could have been made by including testing rates of countries, but these data are not available for most countries. We also do not consider recoveries in our model because we are only analyzing the first weeks of the outbreak. We believe that additional recovery estimates would add unnecessary complexities.

Agradecimentos: This work is partially supported by the project Smart City Concepts in Curitiba (https://smartcityconcepts.chalmers.se/).

References

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore y Piontti, A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M. E., Longini, I. M., and Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science*, 368(6489):395–400.

da F. Costa, L., Rodrigues, F. A., Travieso, G., and Boas, P. R. V. (2007). Characterization of complex networks: A survey of measurements. *Advances in Phys.*, 56(1):167–242.

Dickens, B. L., Koo, J. R., Lim, J. T., Sun, H., Clapham, H. E., Wilder-Smith, A., and Cook, A. R. (2020). Strategies at points of entry to reduce importation risk of COVID-19 cases and re-open travel. *Journal of Travel Medicine*. taaa141.

Gilbert, M., Pullano, G., Pinotti, F., Valdano, E., Poletto, C., Boëlle, P.-Y., D'Ortenzio, E., Yazdanpanah, Y., Eholie, S. P., Altmann, M., Gutierrez, B., Kraemer, M. U. G., and Colizza, V. (2020). Preparedness and vulnerability of african countries against importations of covid-19: a modelling study. *The Lancet*, 395(10227):871 – 877.

Iacus, S. M., Natale, F., Santamaria, C., Spyratos, S., and Vespe, M. (2020). Estimating and projecting air passenger traffic during the covid-19 coronavirus outbreak and its socio-economic impact. *Safety Science*, 129:104791.

IHLG (2019). Industry high level group (ihlg) - aviation benefits report 2019. https://www.icao.int/sustainability/Documents/AVIATION-BENEFITS-2019-web.pdf, accessed in 2020-06-10.

Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Ichii, H., Zacharski, M., Bania, J., and Khosrawipour, T. (2020). The association between international and domestic air traffic and the coronavirus (covid-19) outbreak. *Journal of Microbiology, Immunology and Infection*, 53(3):467 – 472.

Nakamura, H. and Managi, S. (2020). Airport risk of importation and exportation of the covid-19 pandemic. *Transport Policy*, 96:40 – 47.

Sobieralski, J. B. (2020). Covid-19 and airline employment: Insights from historical uncertainty shocks to the industry. *Transport. Research Interdisc. Perspectives*, 5:100123.

World Bank (2020). *Global Economic Prospects, June 2020*. World Bank Group. License: Creative Commons Attribution CC BY 3.0 IGO.

Zhang, R., Li, Y., Zhang, A. L., Wang, Y., and Molina, M. J. (2020). Identifying airborne transmission as the dominant route for the spread of covid-19. *Proceedings of the National Academy of Sciences*.