Using flight data to build a high-frequency global Passenger Capacity Index

**Keywords:** COVID-19, Economic shock, Tourism, High-frequency data, Big Data

# Introduction

COVID-19 related domestic lockdowns and associated restrictions on international travel continue to profoundly impact on economies globally, in respect of employment and output. According to the World Tourism Organization (UNWTO) [1], international tourist arrivals have declined 65% during the first half of 2020, which translates into a loss of 440 million international arrivals and about US$ 460 billion in export revenues from international tourism. Exchange-rate and balance of payments pressures have been experienced by economies heavily dependent on tourism and travel inflows [2]. As economies move at uneven speeds between higher and lower levels of restrictions on travel, policymakers face a shortage of timely official data which further complicates assessment of their economic impacts.

In response, high-frequency data are required to provide a more up-to-date assessment, both nationally and globally, of the economic impacts of tourism and travel restrictions, which will likely persist into 2021 and beyond. In this work we contribute to the assessment by using high-frequency flight data to estimate, with global coverage, changes in international passenger capacity and to compile a Global Passenger Capacity Index (PCI). We discuss how estimates of passenger capacity and the PCI compare with available official estimates of passenger data. We assess the usefulness of our indicators and discuss ongoing work and potential areas for improvement.

# Methods

In this section we describe flight data and how they are collected, we outline our approach to the identification of commercial passenger flights, and we set out the approach for estimating passenger capacity and the development of the Passenger Capacity Index.

## Flight data

Aircraft transponders send a signal which contains information on the aircraft’s location, altitude, etc., and private providers like [FlightRadar24](https://www.flightradar24.com/about) use terrestrial receivers to collect these messages and transform them into structured data feeds. Flightradar24 then combines the real-time information it receives from thousands[[1]](#footnote-1) of aircraft worldwide with flight schedule data from airlines and airports. The term "flight" used here is nonetheless broad and covers commercial flights (including scheduled, passenger and cargo flights) as well as private and hobby flights.

## Enriching FlightRadar24 data

We identify commercial flights by enriching the information provided in the FlightRadar24 data with matched information from departure and arrival airports, airline operators and the aircraft features from several publicly available sources, as follows:

## Airport of departure and arrival: We have derived a subset of airports by matching International Air Transport Association (IATA)\_Airports Codes with an open source database hosted on the [Github](https://github.com/datasets/airport-codes)[[2]](#footnote-3) platform to add information including the airport name and geographical location. Not all airports have an IATA identification code, however airports with scheduled commercial traffic air service will have an IATA code to facilitate booking. The results of the matching exercise provide an airport-to-airport matching structure that can be queried at different levels of detail and geographical aggregates.

* **Airline Operator:** For the airline operator we use a [list of passenger airlines](https://en.wikipedia.org/wiki/List_of_passenger_airlines) extracted from Wikipedia to identify commercial passenger airlines and differentiate this subset in the database. This list includes airlines in operation that offer regular (usually scheduled) services to paying passengers from the general public.
* **Aircraft:** When known, the aircraft model used during the flight in the FlightRadar24 database is identified by the International Civil Aviation Organisation (ICAO) code. We have webscraped aircraft characteristics including the official seat capacity/payload of different aircraft models from a number of databases published online, namely [DOC8643](https://doc8643.com/index), [airwar.ru](http://airwar.ru/) and [airliners.net](https://www.airliners.net/), and converted these unstructured information into a structured data set.

## Estimation of a Passenger Capacity Index

In the absence of a freely accessible data source for passenger numbers with global coverage, we approximate trends in passenger movement by accounting for differing seat capacities of flights. We estimate Passenger Capacity as the average total capacity of an aircraft that can range due to different configurations of seat classes. We compute Total Passenger Capacity for each flight and aggregate successively from the airport-to-airport pairing to derive total level capacity at varying macro-levels including by airport, country, and region.

We estimate the weekly *Passenger Capacity Index* for any given week (Wt) as:

, (1)

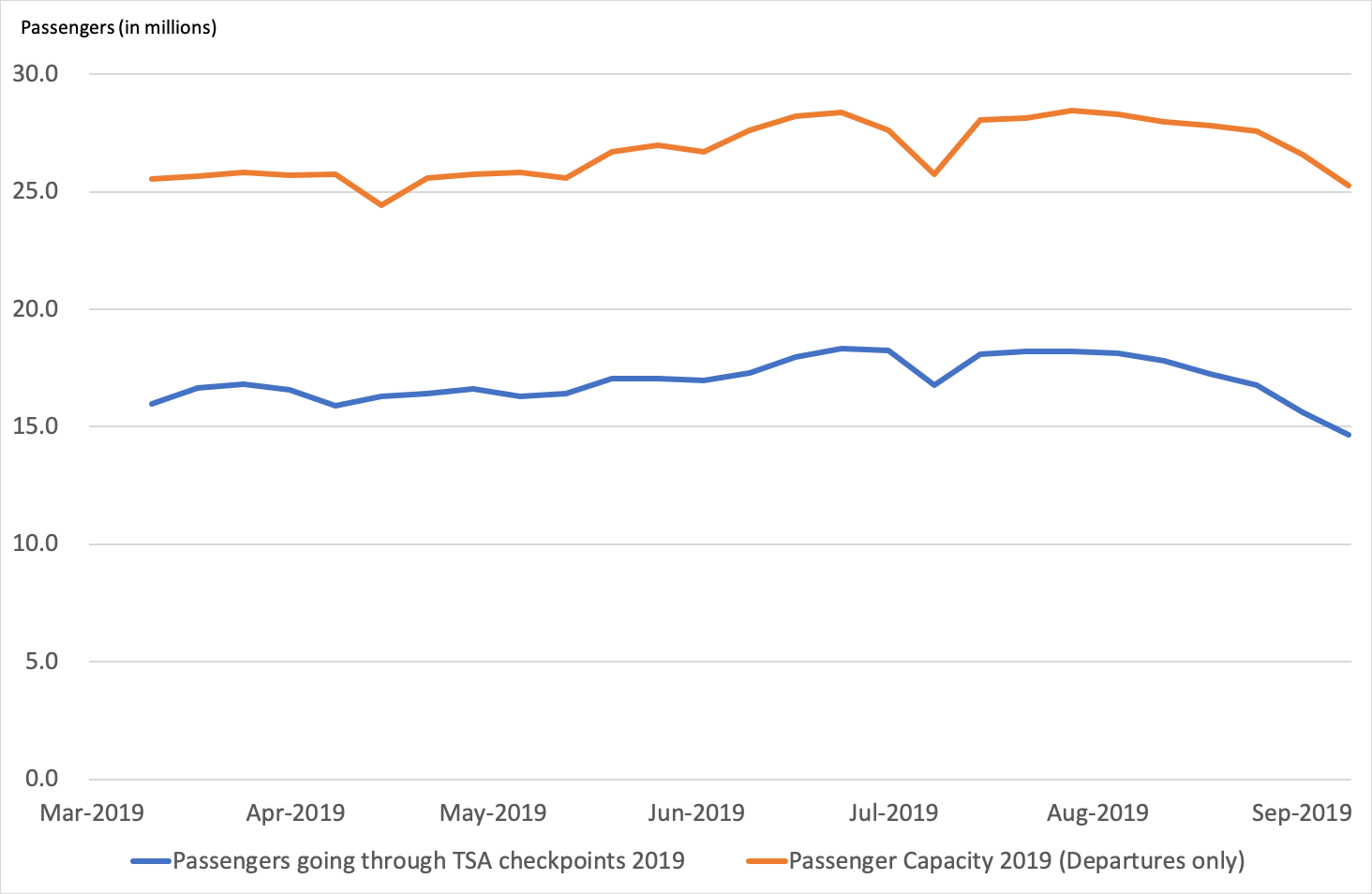
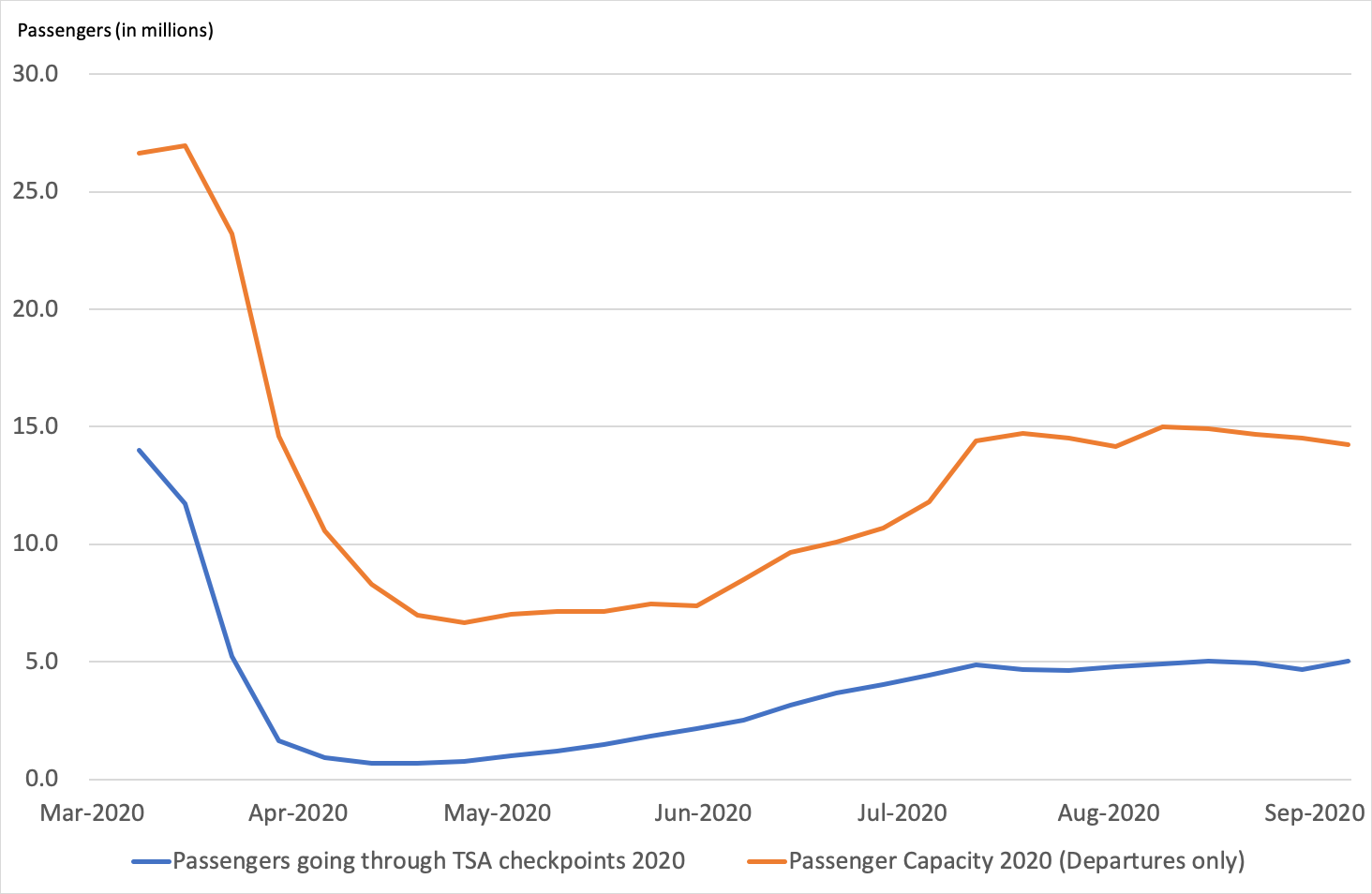
where is the weekly average capacity in 2019. We choose an average of 2019 as the reference period to account for differences in the timing of seasonal fluctuations across countries and regions — some countries might see a surge in flights during summer season while others during winter season.

# Results

In this section, we will first discuss how estimated weekly passenger capacity compares with available official numbers of passengers. We will then highlight some key results and compare trends against other economic indicators.

## Comparisons of estimated weekly passenger capacity against other sources

In Figure 1 we compare our weekly passenger capacity with official number of passengers going through [Transportation Security Administration (TSA) checkpoints](https://www.tsa.gov/coronavirus/passenger-throughput) at US airports[[3]](#footnote-5). We note that passenger capacity is consistently higher than actual number of passengers[[4]](#footnote-6), which is not surprising. In terms of changes in level, while 2019 is fairly similar, the shock caused by the COVID-19 crisis in 2020 suggests the index does not fully take account of the reduced numbers of passengers flying on individual flights due to changes in seats demand and new socially distanced configurations. In 2020, the number of passengers have dropped to as low as 7% in the first week of April when compared to the first week in March and have recovered up to 36% as of September, while passenger capacity dropped to 40% and has recovered to 55% respectively.

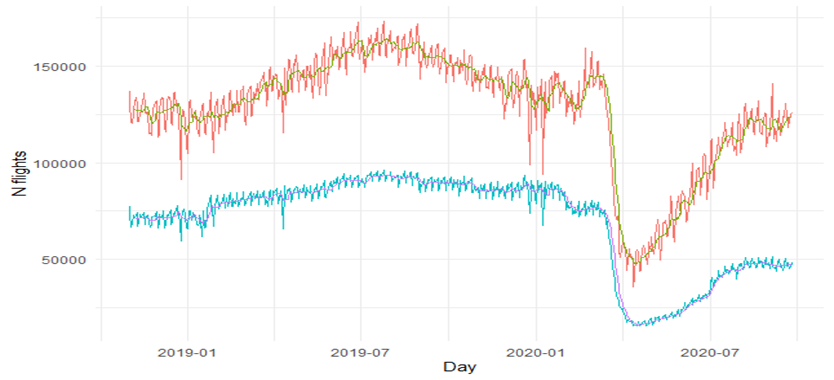
 

1. Data from March to September 2019 (b) Data from March to September 2020

## Figure 1. Comparison of estimated passenger capacity with Transportation Security Administration checkpoints (weekly series).

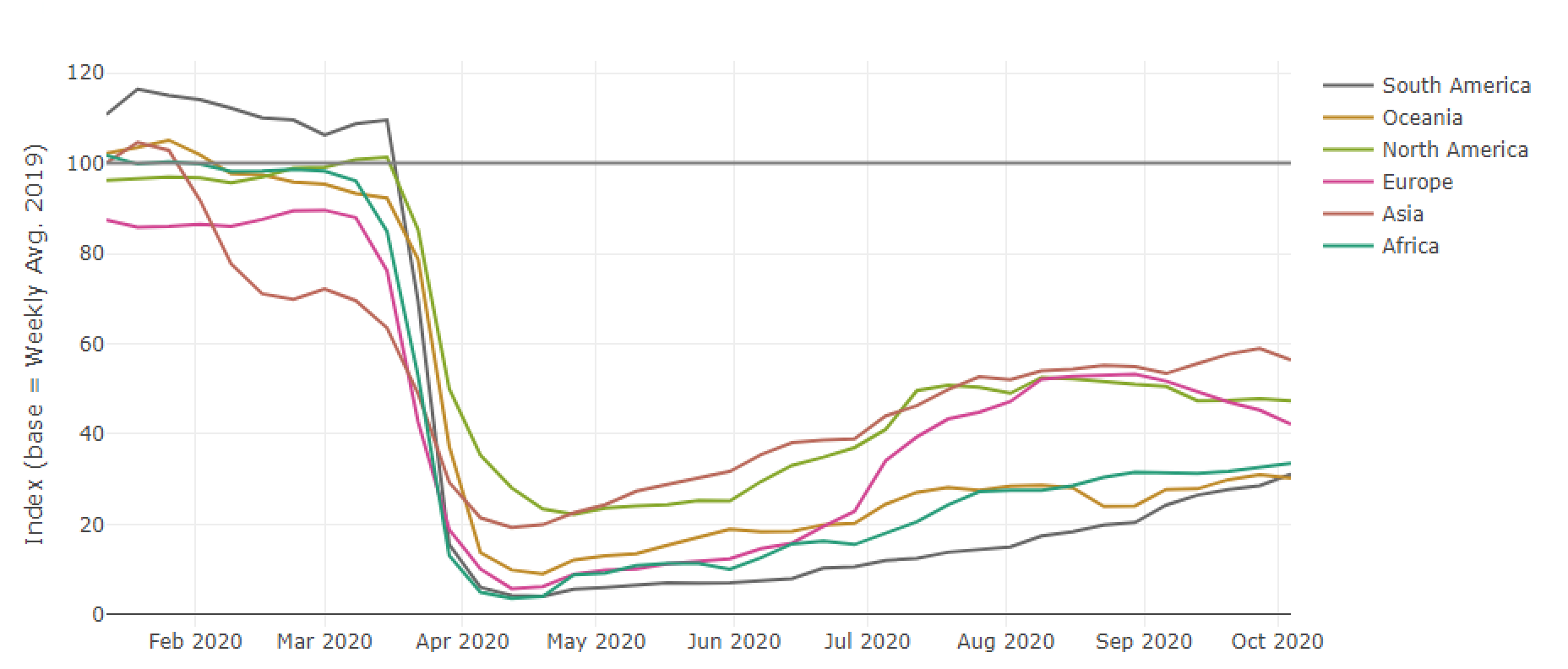
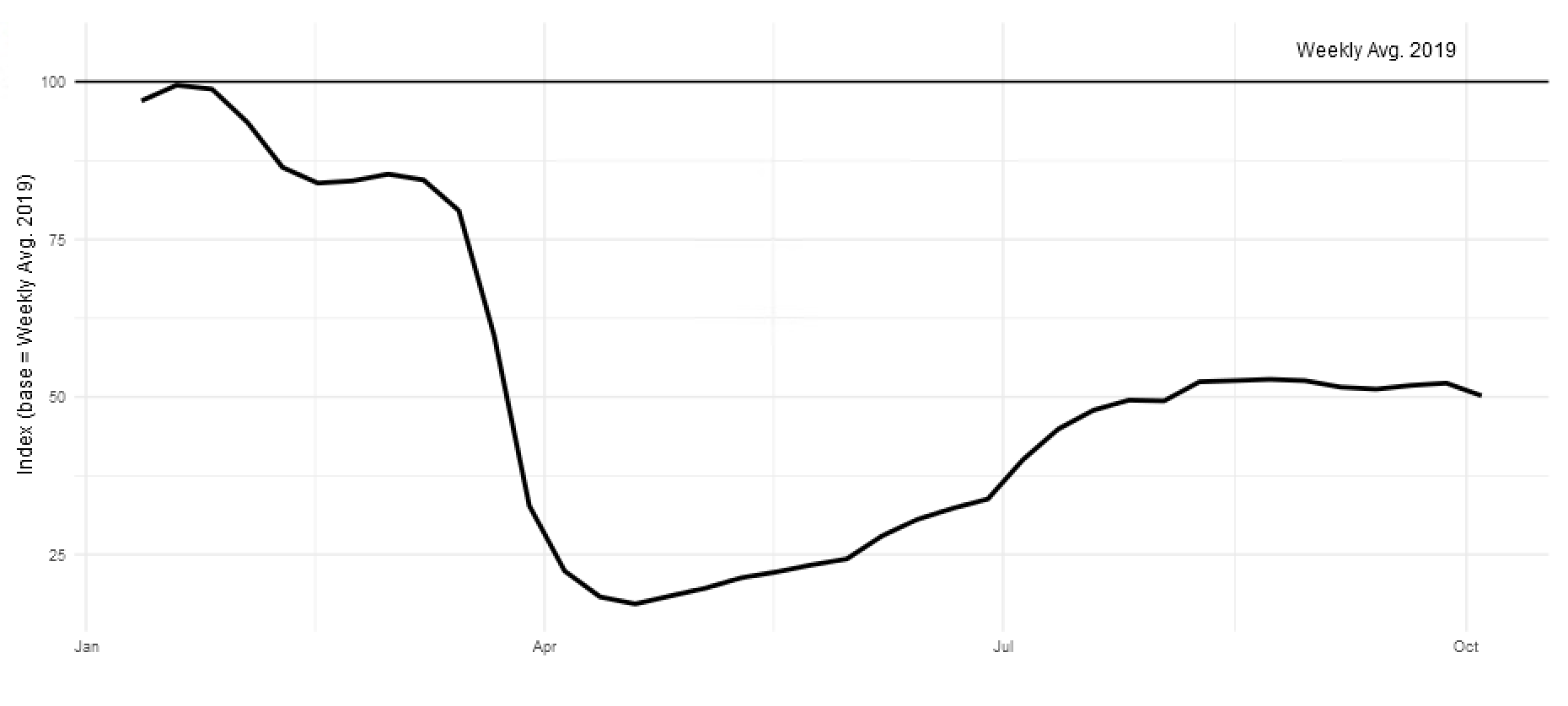
## Key Results

In Figure 2 we compare the daily counts *Total flights* taken from the FlightRadar24 database with our defined series of *Commercial passenger flights* to demonstrate the usefulness of this distinction. The different pace of recovery of the two series reveals a higher reverting behavior in the case of *Total flights*, while that of *Commercial passenger flights* has remained subdued, with levels remaining stalled at a considerably lower level than before the outbreak of the pandemic.



**Figure 2. Daily counts of flights for *Total flights* vs *Commercial passenger flights* only**

The *Regional[[5]](#footnote-7) Capacity Indexes*, reveal very varied degrees of recovery across different regions. Having initially experienced the lowest decline compared to other regions, North America and Asia have gradually increased their passenger capacity. However, Oceania, Africa and South America suffered the largest declines at the onset of the pandemic, and are now recovering at a relatively slower pace. After an initial decline in March, Europe showed signs of relaxed restrictions during the summer which proved to be temporary, and were followed by a second decline from September.

****

**Figure 3. Global Index and Regional Capacity Indexes, January to October 2020.**

## Areas for Improvement

As a further step in our research, we plan to compare Passenger Capacity Index with economic indicators of trade, such as import and exports in services, and Balance of payments data in net revenues from tourism and travel for countries in which these sectors play an important role.

We currently estimate passenger capacity based on the estimated number of seats rather than actual number of people flying. We cannot take account of the reduced numbers of passengers flying on individual flights and new socially distanced configurations. If we can access passenger numbers in the future, we will present those data also.

We cannot at present identify flights operated by passenger airlines as cargo-only flights. We expect that our data shows a temporary increase in such flights during the early stages of the pandemic as passenger airlines mobilized to transport personal protection equipment. We are currently exploring this phenomenon to discern cargo-only flights and create dedicated estimates to provide an overall picture of the commercial industry.

# Conclusions

The effects of the Covid-19 pandemic are devastating for the tourism sector worldwide. Policymakers across the globe require having access to reliable resources of data so that adequate measures can be adopted aiming at mitigating the severe economic consequences in this important sector. In response we demonstrate the usefulness of high-frequency flight data to estimate changes in the number of passenger flights and associated capacity from the level of airport pairing to regional and global totals.

# References

1. UNWTO World Tourism Barometer, Volume 18 Issue 5 (2020), 3.
2. IMF, [Policy Responses to COVID-19 Tracker](https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19), (2020).

1. Across 2019 data, an average of over 180,000 flights were tracked per day. Source: FlightRadar24 and IMF staff calculations. [↑](#footnote-ref-1)
2. GitHub is a code repository hosting service for developers. [↑](#footnote-ref-3)
3. The comparison is restricted to the period between March and September due to limited availability of the TSA data. [↑](#footnote-ref-5)
4. Not all passengers on a plane will go through security checkpoints, for example passengers on domestic connected flights. [↑](#footnote-ref-6)
5. Namely Africa, Asia, Europe, North America, Oceania and South America. [↑](#footnote-ref-7)