Online Travel Booking Site as a Source for Airfare and Lodging Price Indexes

Tashi Edwards

Ben Houck

Craig Brown

FedCASIC Workshop

April 2022

Overview

- Project background

- Data collection

- Price relative calculation

- Research indexes

Project background

- Currently, airfare & lodging prices are manually collected online each month.
  - Airfare: 1,000 quotes every month
  - Lodging: 550 quotes every month

- Economic Assistants collect list prices for each quote by querying companies' websites using fixed specifications.

- Project goal is to investigate if web-scraped prices can replace manual collection.

About Online Travel Booking Site

■ A travel aggregator of consumer travel services:

▶ Airfare

▶ Car rentals

▶ Cruises

▶ Hotel accommodations

Data Collection

■ Each month, we generate an input file based on the CPI sample of quotes that is used as search criteria to query the travel booking website.

▶ Only 66% of the airline sample is collected in research.

▶ Around 85-90% of the lodging sample is collected through web-scraping our data source.

Data Collection

CPI Airfare Input File

- Origin city code
- Destination city code
- Fare class
- Departure/return dates
- Pricing period

CPI Lodging Input File

- Query ID
- Pricing period
- Unique property ID
- Check In/Out dates
- Number of guests

Data Collection

Frequency and timing of data collection:

- Pricing Period (PP) is 9 days long; 3 pricing periods/month
- Lodging: Each query is collected 1x per PP
- Airfare: Each query is collected 3x per PP

Lodging data elements collected:

- ▶ Property characteristics/amenities
- ▶ Room ID and Descriptions
- ▶ Payment Model/Cancellation Policy

Airfare data elements collected:

- ▶ Origin/destination airport codes
- ▶ Airline name/carrier code
- ▶ Number of layovers
- ▶ Fare class/codes for each flight leg
- ▶ Departure/return time

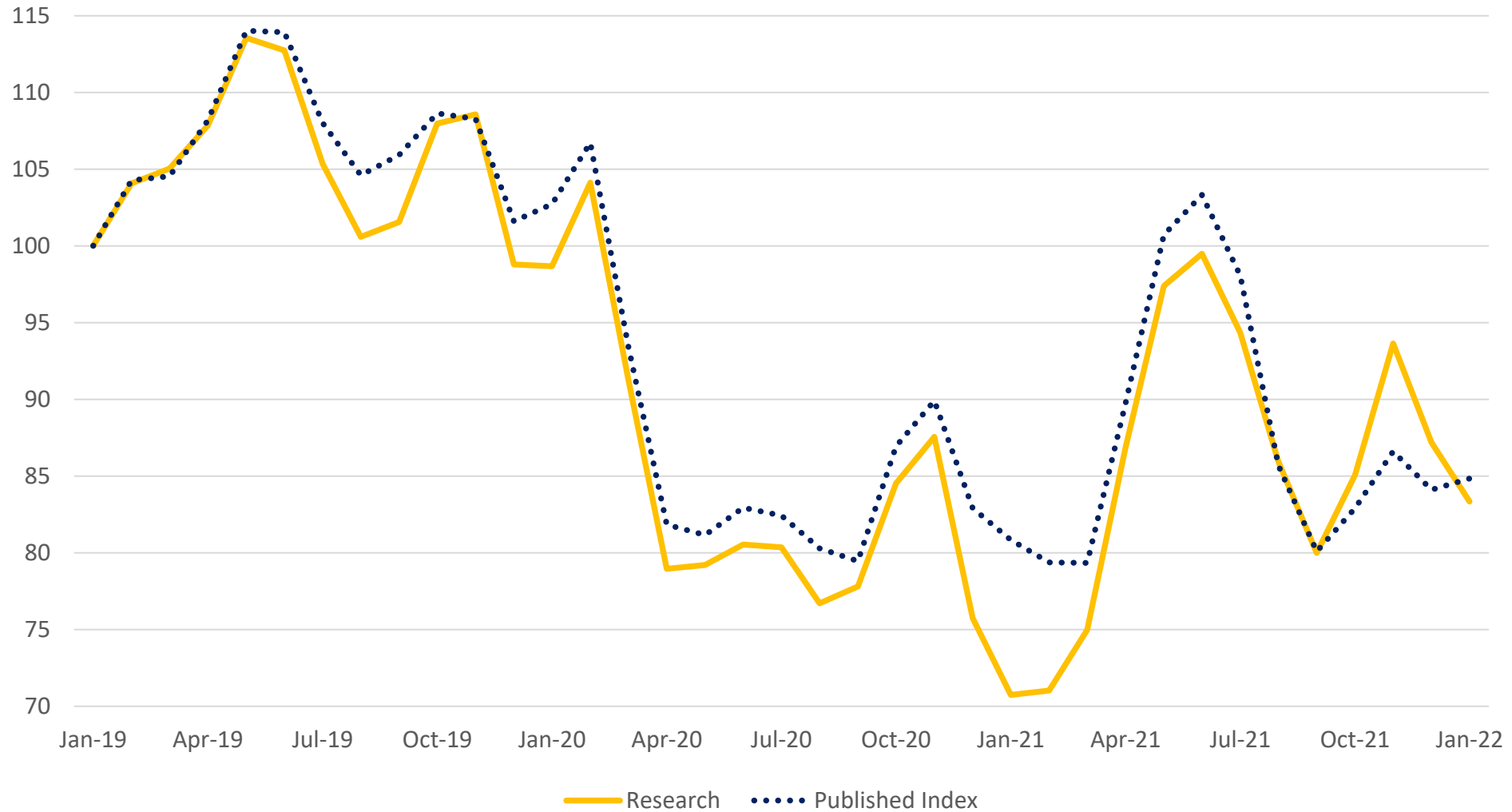# Price Relative Calculation (Airfare) (1)

1. Matching reservations

   ▶ Create unique combinations of collected data for each reservation

   ▶ Airline code/origin airport code/destination airport code/ trip type/departure date/return date/fare class

2. Quote overlay

   ▶ Replace the CPI price with the web-scraped price for each matched quote

Price Relative Calculation (Airfare) (2)

■ On average,7% of CPI quotes priced through this outlet could not be replaced using web-scraped data.

▶ Reasons for non-matches include:

– Seasonality of airfare

– Airline service changes

– Certain fare characteristics were unavailable

– Coronavirus pandemic

■ Quotes we did not attempt to collect through web-scraping are included using their CPI collected prices.

BLS

Research Index (Airfare)

Legend: Research (solid line), Published Index (dotted line)

Price Relative Calculation (Lodging) (1)

1. Match reservations

   ▶ Create unique combinations of collected data for each reservation

   ▶ Match Quote/PP/Qry Num/RoomID/Payment/Cancel

2. Calculate reservation relatives (current period price/previous period price)

3. Calculate quote relatives

   ▶ Geometric average of reservation relatives for each quote

Price Relative Calculation (Lodging) (2)

- Experiment began in November 2019 and we have results through October 2021.

- Each month since September 2020, we've:
  - Sent an average of 24 queries per quote for collection
  - Collected an average of 203,000 reservations
  - Matched an average of 108,000 reservations
    - Average of 226 reservations per quote

- Instances of non-response dropped 40% over the period.

Research Indexes (1) (Lodging)

- Use price relatives we calculated from web-scraped data for quotes with usable reservations.

- Quotes that we attempted to collect data for, but where no matches were found, are imputed.

- Quotes we did not attempt to collect through web scraping are included using traditionally-collected prices.

Research Indexes (2) (Lodging)

Legend: Geometric Average — Published Index (dotted)
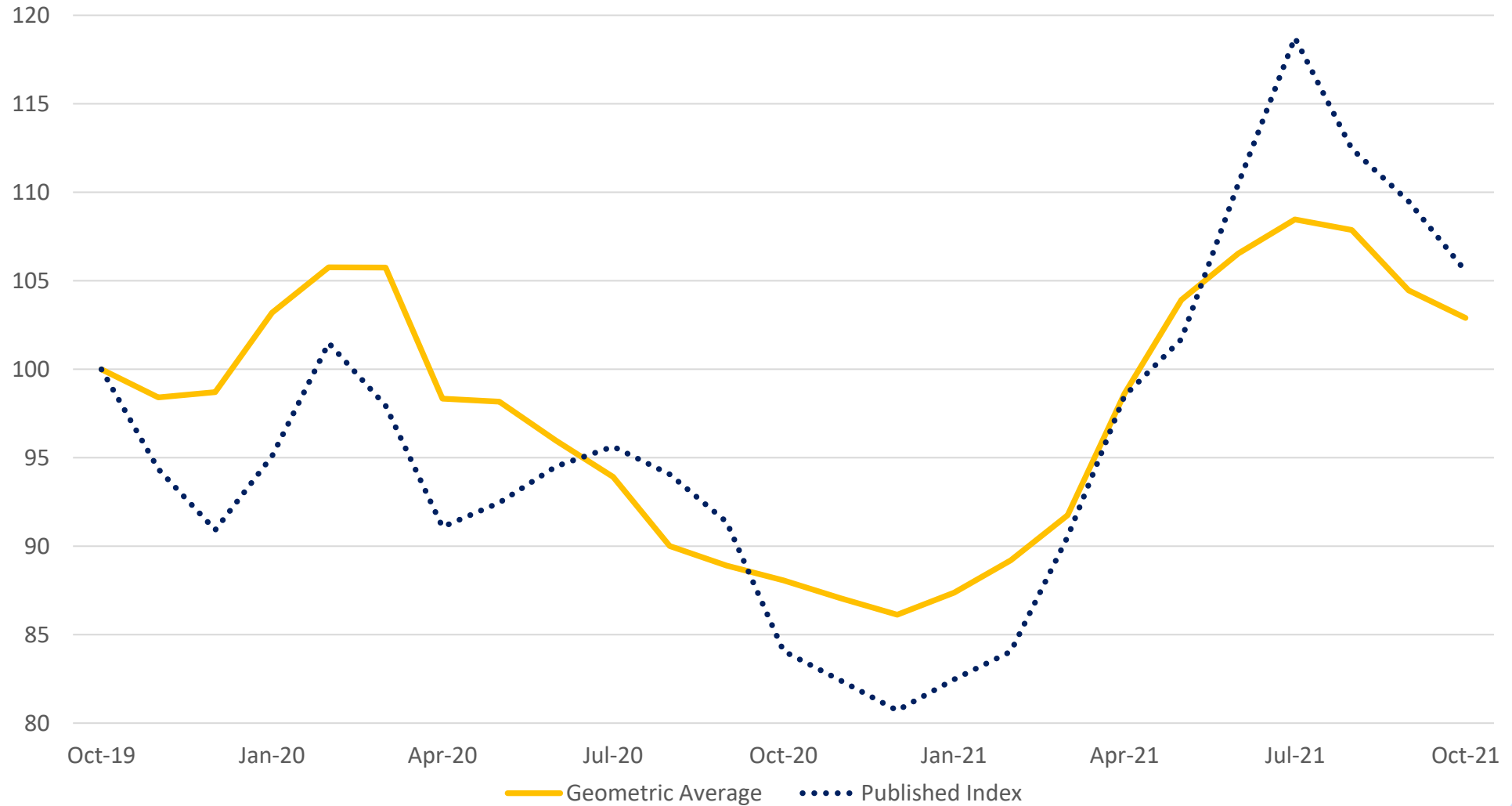
Summary

- Carefully consider your data requirements and pros and cons of various sources.

- Communicate your requirements to your source and get permission to web-collect.

- Determine if you can use an Application Programming Interface (API).

Authors/Presenters

Tashi Edwards

edwards.tashi@bls.gov

Ben Houck

houck.ben@bls.gov

Craig Brown

brown.craig@bls.gov

Economists, Consumer Price Index
U.S. Bureau of Labor Statistics