Measuring inflation of ride sharing services in Brazilian official CPI

Online UNECE meeting of the Group of Experts on Consumer Price Indices
June 2021

Vladimir G. Miranda, Lincoln T. da Silva, André F. G. Almeida, Pedro K. da Costa

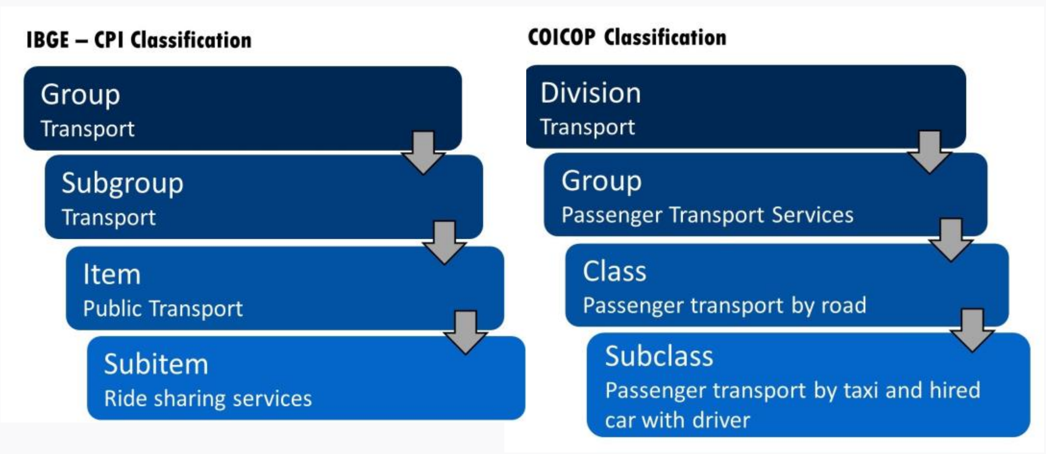
vladimir.miranda@ibge.gov.br

Brazilian Institute of Geography and Statistics – IBGE
Price Indices Coordination - Directorate of Surveys

Origin of the issue

As a result of the last HBS update, expenditures on ride sharing services were observed and have shown to be representative.

Illustration of how such expenditures were classified



Comparison of final weights observed, for each area, for taxi and ride sharing services.

Area	IPCA Taxi	IPCA Ride sharing Services	INPC Taxi	INPC Ride sharing Services
BR	**0,21**	**0,21**	**0,16**	**0,15**
AC	0,54	-	0,55	0,07
PA	0,43	-	0,32	-
MA	0,32	0,11	0,41	0,15
CE	0,18	0,15	0,15	0,16
PE	0,30	0,32	0,15	0,28
SE	0,58	0,11	0,53	0,17
BA	0,38	0,30	0,19	0,21
MG	0,24	0,19	0,17	0,16
ES	0,12	0,10	-	0,09
RJ	0,45	0,31	0,20	0,26
SP	0,16	0,20	0,11	0,12
RS	0,26	0,38	0,20	0,27
MS	0,09	0,23	-	0,28
GO	-	0,26	-	0,09
DF	-	0,25	0,11	0,16

Origin of the issue

We were faced with these results in mid 2019 and had to come with a solution to implement the component in the basket update in January 2020.

Some questions we had to deal with:

How to measure it?

How to define a given service to track the price and fit the matched model method?

Is it possible to derive local indicators?

How to get access to data that fit our needs?

Main components of the service price

In order to answer these questions, it is key to look at how a given ride is priced.

Price of a trip is composed of three main components:

Base rates, Booking fees and Surge multiplier

"Rigid" components

Bare prices: per km and per minute rates applied to a trip.

Booking prices: flat fee based to support operational, regulatory, and safety costs.

"Flexible" component

Surge multipliers: a dynamic factor that is charged according the balance between riders demand and drivers supply in a given area and time.

How to track price changes?

Different approaches can be derived to try to capture the price evolution. For instance, by tracking the price of a "standard" trip fare based only on the "rigid" components.

This is similar to what is usually performed to measure CPIs for taxis.

However, ride sharing services have remarkable pricing dynamics and this approach may fail to capture this. Also, geographical nuances might not be captured appropriately.

On the contrary, such services resemble price dynamics of sectors such as airfares where prices rely on seats availability.

In such cases, the CPI manual suggests an approach based on following a representative number of trips:

"The index for airline and other public transport services should use the prices of a sample of specific trips rather than revenue per kilometer or per passenger-kilometer. If the CPI has strata for different geographic locations, that is typical in large countries, points of origin (e.g., airports, train stations, or motor coach stops) should be chosen in each location, and trips selected within origins or destinations in those locations. There are generally multiple classes of service. In addition, transport fares may vary by day of the week, time of day, time of year in response to variations in demand, and how long in advance the ticket is purchased. The selected trips should reflect this variety but hold these variables constant." CPI manual Concepts and methods, draft January 2020, paragraph 11.289.

We here try to develop an approach based on this idea.

How to get the prices?

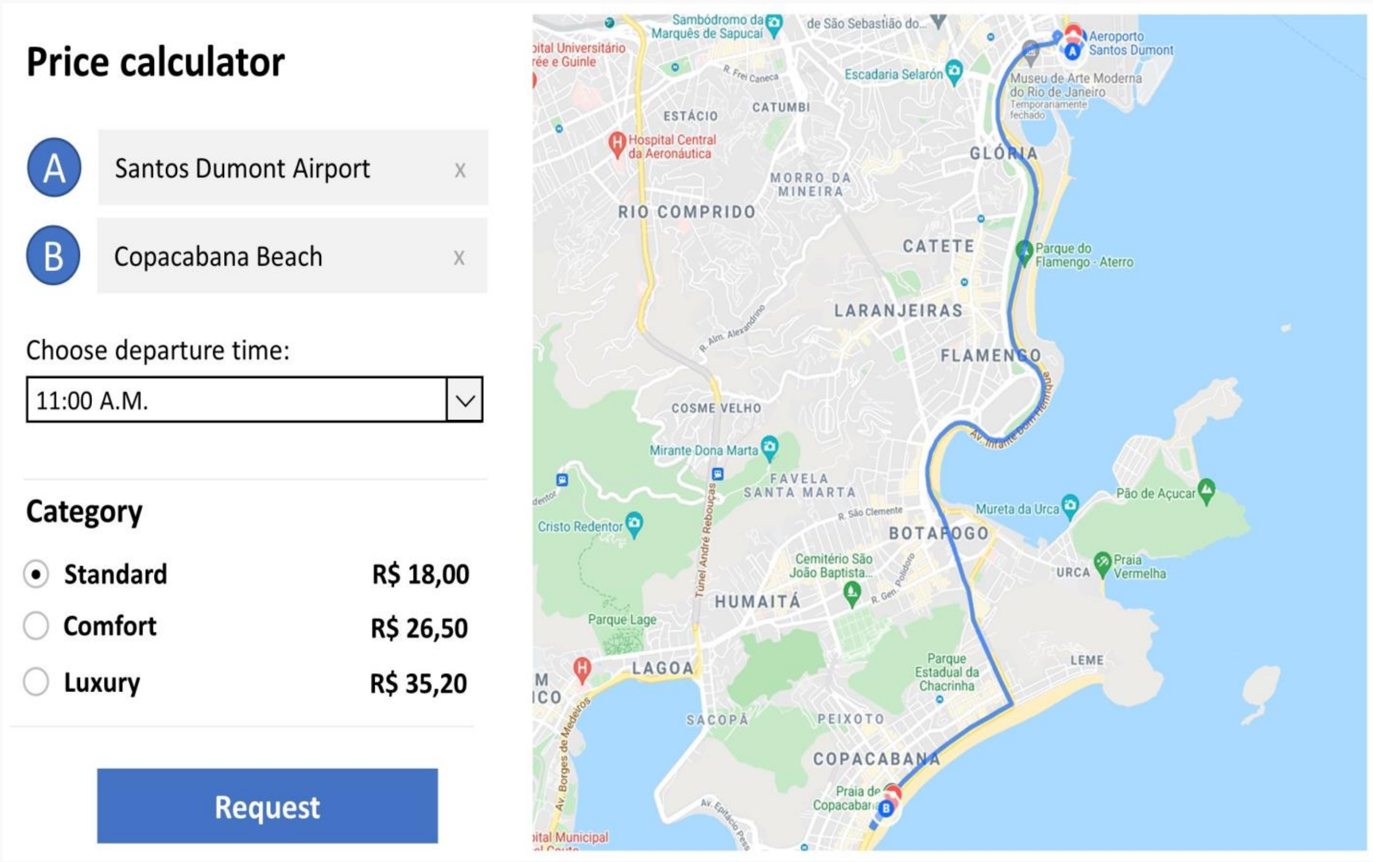
Ideally: transactions data information.

As we don't have access to these data sets, web can provide a good opportunity.

Price simulator sites can provide the desired data. A price is derived after some input parameters defining a trip are provided, similar to a user demanding a ride via its mobile app.

Prices may differ from what is seen by a given user, though they are able to provide realistic estimates and capture all price components as shown in the results.

This offers a good opportunity since prices can be collected via web scraping techniques.



How to define the trips to track?

Analysis of the price simulator's sites show that a trip is defined by a set of input parameters, namely: departure and arrival places, departure time, category of the service, company.

Without transaction data to define most representative trips, in principle an infinite combination of routes and departure times are available.

As our initial proposal we defined a small number of trips to track based on some arbitrary choices:

Number of routes: 5 different routes were selected for each area.

Departure places: encompasses places with high demand for this sort of service, such as shopping centers, airports, markets, and touristic spots.

Typical distances: trips with average distances of 3, 5 and 10 km were selected. Such distances are in agreement with most typical taxi rides.

Arrival places: set after manual inspection based on departure places and typical distances.

Category: only the standard category is considered.

How to define the trips to track?

Departure time: for each route, five different departure times are considered.

Based on this, an example of a specific trip would be:

Santos Dumont Airport – Copacabana Beach, standard rate, 11 A.M, company X.

Frequency of collection: For each trip, prices are collected daily for weekdays of the month.

This already amounts to collecting around 500 prices per area and over 7000 prices for the national aggregate each month.

Collection method

Prices are collected automatically via web scraping by home-made scrapers using R and RSelenium package.

The robot mimics a human inserting the inputs in the price simulator and extracting the resulting price.

The collection is scheduled to occur near the trip departure times predefined as this is the most common practice adopted by consumers.

The restriction on the number of prices and days of the week amounts to the fact that if the robots have any problem a manual collection can be performed as a fallback plan.

Index methods

Methods used follow an approach similar to the one adopted for other subitems of the basket.

1) monthly average per trip:

$$\bar{p}_{\alpha}^{l,t} = \frac{p_{\alpha,i}^{l,t}}{n_{\alpha}^{t}}$$

2) Relative of mean prices per trip for following months:

$$R_{\alpha}^{l,t-1:t} = \frac{\bar{p}_{\alpha}^{l,t}}{\bar{p}_{\alpha}^{l,t-1}}$$

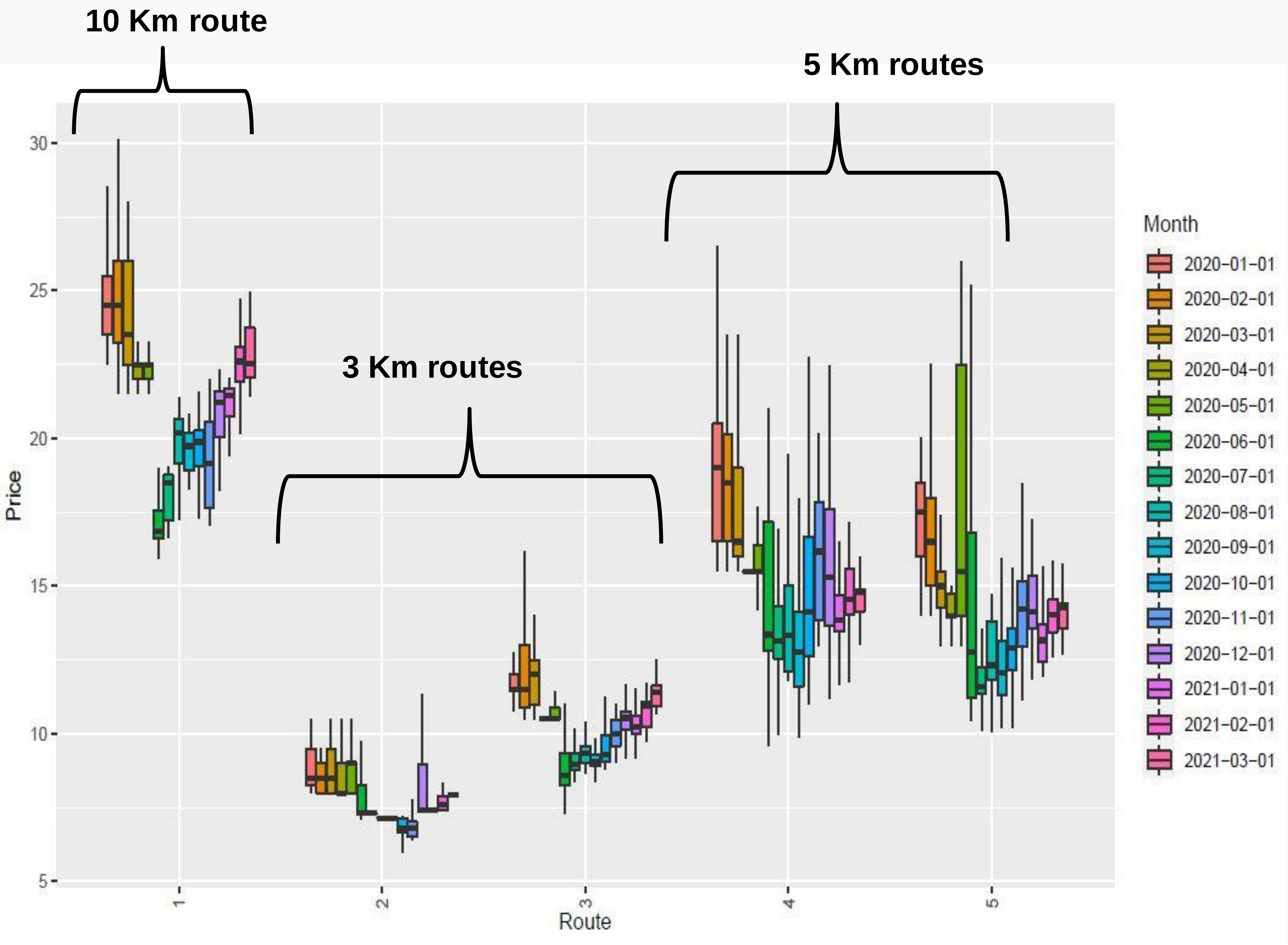
3) Areas indexes: different trips results are aggregated via geometric average

$$I^{l,t-1:t} = \prod_{\alpha} \left(R_{\alpha}^{l,t-1:t}\right)^{1/N_{\alpha}^{l}}$$

4) National index:

$$I^{BR,t-1:t} = \frac{\sum_{l} W^{l} w^{l,t-1} I_{\alpha}^{l,t-1:t}}{\sum_{l} W^{l} w^{l,t-1}}$$

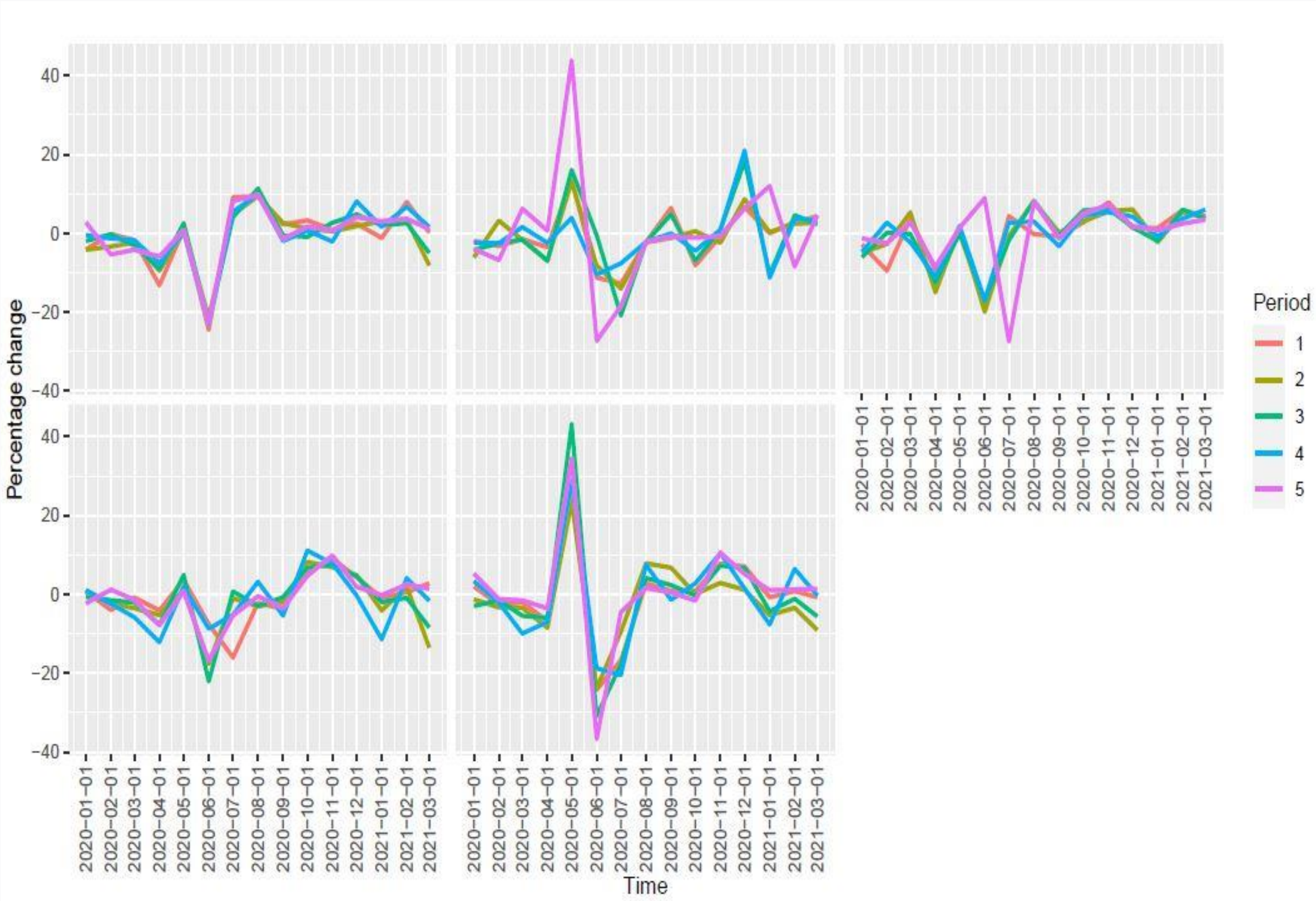
Monthly prices distribution for São Paulo



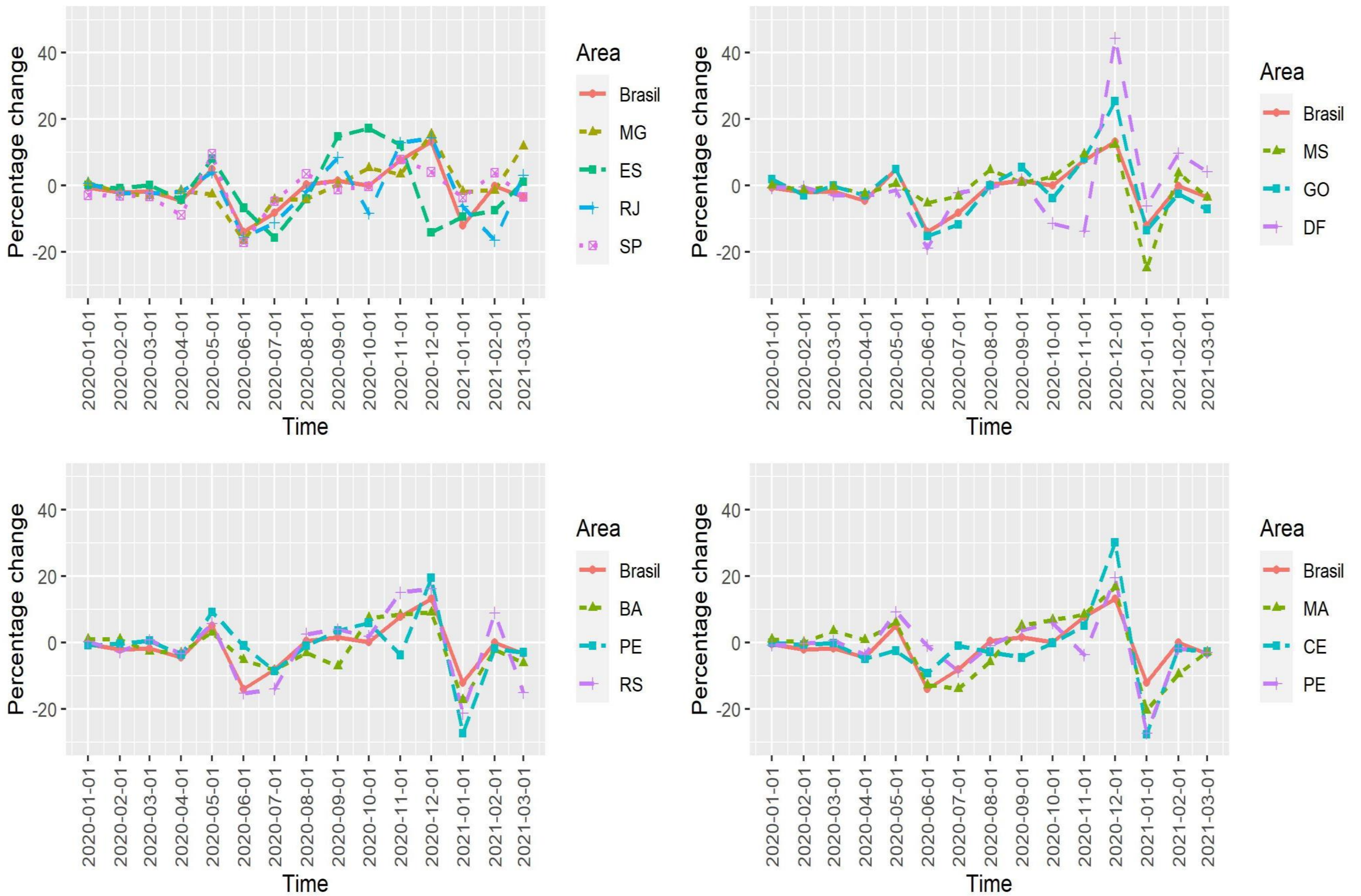
IBGE

IBGE Instituto Brasileiro de Geografia e Estatística

Price change (%) series for São Paulo



IPCA Price changes (%) series per area and nationally



Final remarks

The approach developed based on the automatic collection of web prices from price simulator sites has shown to be able to provide price estimates that are able to get the price dynamics of the sector.

The consistency in the results at different regions, departure times and aggregation levels suggests that the method is able to provide robust results even with a reduced number of trips.

Use of web scraping techniques is able to extract a significative number of prices at timely and efficient manner avoiding an extra burden of the monthly collection of over 7000 prices by the field staff.

The method is able to provide results for different areas hence allowing to track the peculiarities of the price dynamics locally.

The method can be improved by the selection of more representative routes. More automatic and refined techniques to select the routes can be devised. However, caution should be taken in the selection of the routes. A manual inspection might be necessary.

With an automatic way to select the routes a massive approach to collect the prices can be tried. However, this is more computationally resource-intensive and more prone to IP blocks. In case of failure, manual collection will not be able to reproduce the robots' collection.

Thank you for your attention.

Measuring inflation of ride sharing services in Brazilian official CPI

Online UNECE meeting of the Group of Experts on Consumer Price Indices
June 2021

Vladimir G. Miranda, Lincoln T. da Silva, André F. G. Almeida, Pedro K. da Costa

vladimir.miranda@ibge.gov.br

Brazilian Institute of Geography and Statistics – IBGE
Price Indices Coordination - Directorate of Surveys