Aggregate Time Series Analysis of Urban Bus Transportation Demand in São Paulo City

Celio Daroncho
Phd student, UNICAMP, Brasil
Celio.daroncho@fatec.sp.gov.br

Pedro José Perez Martinez
Professor PhD, FECFAU-UNICAMP, Brasil
pjperez@unicamp.br

Received: March 29, 2024
Accepted: July 15, 2024
Online Published: August 7, 2024
ADSTRACT
This study aims to analyze the impact of the COVID-19 pandemic on urban bus transportation demand in São Paulo, Brazil. Using time series data on passenger ridership from 2005 to 2023, we divided the dataset into two periods: pre-pandemic (2005-2020) for model generation and post-pandemic (2020-2023) for comparison of forecasted and observed values. We applied the Seasonal Autoregressive Integrated Moving Average (SARIMA) method using Jamovi software and the R programming language to analyze the time series and generate a forecasting model. Our results indicate an average forecast of 206 million passengers per month in 2023, compared to an observed average of 173 million passengers per month—a 19% shortfall. Furthermore, this forecast remains 21% below the 2019 average (220 million passengers per month). The annual projected decline between 2011 and 2023 is approximately 15.8%, resulting in a loss of over 35 million passengers annually. Notably, the decline in passenger volume predates the pandemic, which exacerbated the situation. Recovery prospects remain uncertain and hinge on factors such as sectoral investment realignment and incentives for urban bus system usage.


1 INTRODUCTION

Urban mobility constitutes an essential service to ensure people’s access to the city and citizenship (Rodrigue, 2020) and is recognized as a social right by Brazilian legislation (Brasil, 2012). Nevertheless, there is an observed downward trend in demand for public transportation, influenced by economic factors, shifts in travel patterns, and, more recently, the effects of COVID-19. This underscores the need for a deeper analysis of the factors contributing to the decline in demand for this service (Cardozo et al., 2023; Wagner and Marujo, 2023; Faria et al., 2023).

The global spread of SARS-CoV-2, first identified in Wuhan, China, at the end of 2019 (WHO, 2024), profoundly impacted both global health and urban mobility. Social distancing measures led to significant changes in travel patterns, including a substantial reduction in public transportation usage and an increase in remote activities (Daroncho and Martinez, 2024). However, a detailed understanding of the pandemic’s implications for Urban Public Transportation (UPT) demand still lacks robust statistical analyses, particularly regarding permanent transformations in population travel behavior. São Paulo, along with other major global cities, faces challenges related to low bus transportation demand—a trend that predates the pandemic (Daroncho and Martinez, 2024).

2 COVID-19 PANDEMIC

Shortly before the close of 2019, a new disease of catastrophic proportions emerged in Wuhan, China—the SARS-CoV-2 virus, commonly known as Coronavirus—which led to the COVID-19 pandemic (WHO, 2024). The disease rapidly spread worldwide, facilitated by modern transportation’s ease of mobility and extensive interconnectivity. On March 11, 2020, the World Health Organization (WHO) declared a Public Health Emergency of International Concern (PHEIC) or a pandemic, which persisted until May 4, 2023 (WHO, 2024).

In Brazil, pandemic responses were decentralized, with states and municipalities independently implementing quarantines based on local needs (Moraes, 2020). São Paulo State declared a quarantine on March 23, 2020, which underwent several extensions and varying
levels of restrictions, ultimately ending on November 1, 2021 (São Paulo, 2024). According to Moura et al. (2022), Brazil experienced three major COVID-19 waves: the first from February 23 to July 25, 2020; the second from November 8, 2020, to April 10, 2021; and the third from December 26, 2021, to May 21, 2022. In the Metropolitan Region of São Paulo (MRSP), these waves resulted in over 2 million cases and 80,000 deaths by the end of 2022, as reported by the Ministry of Health (Brasil, 2024).

Amid this pandemic context, Feroze (2020) highlighted the need to rethink transportation policies. Pre-pandemic, the focus was on demand management, intelligent technological interventions, and sustainable mobility. However, the public health crisis urgently necessitated reconsideration of transportation’s role in post-COVID economic recovery. This environment provides an opportunity to reshape policies and practices related to urban mobility (Feroze, 2020).

3 URBAN MOBILITY BEYOND THE PANDEMIC

The challenge of declining demand for UPT, affecting many major cities worldwide, predates the COVID-19 pandemic but was exacerbated by it when demand plummeted. Kantar (2021), who analyzed mobility challenges in 32 global cities in 2019, highlighted the significant impact of remote work on transportation habits. This resulted in reduced use of public and shared transport, accompanied by an increased preference for healthier modes such as walking and cycling. TSC (2020), examining data from over 150 metro systems worldwide, revealed that the sector suffered severe pandemic-related effects, with demand dropping by approximately 75%. Furthermore, expectations suggest that post-pandemic demand will remain significantly below pre-pandemic levels.

NTU (2022) had already observed passenger demand decline over the years, emphasizing the cost-revenue gap faced by public transportation systems even before the pandemic. The drastic revenue reduction caused by the pandemic further exacerbated this delicate situation for bus-based UPT in Brazil. Quintella and Sucena (2020) underscored the critical role of public transport in city functioning and demonstrated that passenger volume reduction in Brazilian urban bus systems occurred prior to the pandemic. Factors contributing to this decline—both before and during the pandemic—include economic crises, increased remote work, and inadequate transportation infrastructure investment. Daroncho and Martinez (2024) further substantiated that the issue of public transportation demand, affecting most cities globally, predates the COVID-19 pandemic, with UPT losing passengers since the 2010s.

3.1 Urban mobility in the pandemic

Since the onset of the pandemic, numerous studies have examined the impact on urban mobility and UPT. Aloi et al. (2020), analyzing data from traffic counters, public transport systems, and traffic control cameras in Santander (Spain), observed a 76% reduction in urban mobility. Notably, this reduction coincided with a significant decrease in public transport usage, NO2 emissions, and traffic accidents. Fatmi (2020) investigated travel patterns in British Columbia, Canada, revealing a reduction of over 50% in local and regional trips. Interestingly,
work-related travel increased for specific occupations, while telecommuting became more prevalent among higher-income groups. Additionally, most long-distance trips shifted to private car travel.

Kraemer et al. (2020) demonstrated that travel control measures and restrictions implemented in China significantly curtailed the spread of COVID-19. These measures were particularly effective in mitigating case importation from Wuhan and controlling local transmission. Budd and Ison (2020) emphasized the need to rethink transportation policies and practices. They argued that the crisis presented an opportunity to redefine more sustainable post-COVID transport policies. Notably, this period witnessed a substantial reduction in emissions and spatial shifts in urban travel patterns. Abdullah et al. (2021) investigated travel behavior in Pakistan, observing an increase in shopping-related trips and a general decline in non-work-related travel. They also noted a rise in non-motorized modes for short distances and increased private car usage for longer trips.

As evident from the data, passenger demand in Brazilian UPT was already declining before the pandemic (NTU, 2022). During the initial pandemic period (March to June 2020),
demand experienced a considerable drop, gradually recovering after June 2020 (Daroncho and Martinez, 2024). Amid this demand recovery process, a critical question arises: What is the ideal post-pandemic demand volume? Several studies have addressed this by forecasting post-pandemic demand based on pre-pandemic data.

Cardozo et al. (2023) analyzed UPT demand in the Recife Metropolitan Region during the pandemic, using time series analysis and ARIMA modeling with data spanning January 2007 to February 2020. Results indicated that the pandemic exacerbated the existing downward trend in demand, even beyond the critical period. The forecast for 2020 to 2022 revealed an average loss of 15.36% in demand compared to expectations. Faria et al. (2023) examined collective transport demand in the Goiânia metropolitan region, considering data from January 2009 to November 2022. Using Bayesian Structural Time Series (BSTS) modeling, they found a substantial 53% reduction in actual trips due to the pandemic. Marujo and Wagner (2023) analyzed bus-based public transportation demand in Rio de Janeiro, using data from March 2020 to December 2022 and ARIMA modeling. Their results revealed a 29.5% decrease in transported passengers during the post-pandemic period compared to the pre-pandemic period.

These findings underscore the need for measures to reverse this trend and ensure UPT sustainability in the region. They also highlight the necessity of considering new strategies to make public transportation more attractive to passengers after the pandemic. Moreover, they indicate a shift in user behavior within the UPT system.

4 STUDY AREA AND TRANSPORTED PASSENGER DATA

This study examined passenger data from the municipal bus system in São Paulo City, the capital of São Paulo state in Brazil (Figure 1). With a population exceeding 12 million inhabitants across an area of over 1,500 km², São Paulo ranks as the fifth most populous city globally (IBGE, 2023). The São Paulo City is served by a dense bus network operated by 38 public transportation concessionaires, comprising 1,347 bus lines and 11,925 vehicles. The system is managed by São Paulo Transporte S/A - SPTrans, a municipal mixed-economy public company (SPTrans, 2024).

The history of São Paulo’s bus transportation system includes several reorganizations. In 1991, the system was municipalized through Law 11037 (May 1991) and Municipal Decree 29.945 (July 1991), initiating a process of prioritizing public transportation (Whately and Néspoli, 2013). During this period, from 1984 to 1996, the system transported approximately 1.9 billion passengers annually (SEADE, 2024).
In 1995, the Companhia Municipal de Transportes Coletivos¹ (CMTC), which had operated the system since 1946, was privatized, leading to the creation of SPTrans—an agency responsible for managing the system through concessions. During this period, there was a significant rise in informal or clandestine transportation, causing regular public transport to lose passengers—many of whom were not accounted for in the official fare collection system (Hirata, 2012). Annual passenger numbers declined from 1.6 billion in 1997 to approximately 1.0 billion in 2001 and 2002 (SEADE, 2024).

In early 2004, the Bilhete Único (BU)² system was introduced for SPTrans-managed bus services, allowing unlimited trips within a 2-hour window. By 2005, it expanded to 4 trips within 2 hours, and in December 2005, the BU system was integrated with rail-based transportation (PMSP, 2014).

Figure 2 illustrates the annual demand and its variation for buses. Notably, the period from 1979 to 1996 showed relative stability, followed by a decline from 1996 to 2003. Demand recovered in 2004 and shifted significantly from 2005 onward, with growth until 2011 and subsequent declines from 2011 to 2019—particularly sharp from 2018 to 2019. Interestingly, the demand levels in 2022 and 2023 are nearly equivalent to those observed in 1986.

---

¹ Municipal Company of Collective Transport.
² A card for use in the bus system of the São Paulo City.
For this study, we analyzed monthly data spanning from 2005 to 2023, sourced from the São Paulo City Hall’s information portal (PMSP, 2024). Figure 3 illustrates the annual behavior of the data during this period. Notably, the annual passenger volume experienced growth until 2011 (reaching a peak) and remained relatively stable until 2016. However, a sharper decline occurred in 2015, followed by a partial recovery in 2016. Subsequently, from 2017 to 2019, there was another pronounced decline. In 2019, the passenger volume only exceeded the 2005 level, remaining 10.3% below the 2011 peak. The year 2020 witnessed an abrupt drop due to the impact of COVID-19, followed by gradual recovery starting in 2021. By 2023, the volume remained 17.1% below the 2005 level and 29.3% below the 2011 peak. Notably, the 2023 volume was 21.2% lower than that of 2019—the second-lowest value in the series—with a mere 1.4% growth compared to 2022.

These data alone reveal that the bus system was already losing passengers. These passengers may have shifted to rail-based transportation or individual modes. Additionally, changes in social habits—such as working closer to home or telecommuting—could also contribute to this trend.
4.1 Monthly Transported Passenger Data

In Figure 4, the seasonal patterns of monthly transported passenger data from 2005 to 2013 are evident. Interestingly, starting in 2022, the monthly seasonality reemerged in the data.

Figure 4: Monthly Commute by São Paulo City’s Urban Bus System (2005–2023).
The vertical dashed red line represents the onset of the COVID-19 period.

The seasonal behavior of travel data, as depicted in Figure 5, reveals month-to-month variations. From 2005 to 2019, except for January and February, the data consistently fell within the range of 200 to 270 million trips per month. This pattern exhibited a monthly seasonality, characterized by peaks in March, May, August, and October. Notably, the monthly behavior remained remarkably similar across all years, except for 2019. In that year, March deviated partially from the established pattern observed in previous years. However, this deviation was rectified in 2022 and 2023, as the typical seasonal trend reemerged. Furthermore, a close examination of the monthly volumes in 2022 and 2023 reveals their striking similarity. While 2023 exhibits relatively higher values in January, March, and June, it experiences a relatively lower volume in November.

Figure 5. Seasonality of Monthly Commute by São Paulo City’s Urban Bus System (2005–2023).
The gray shaded region represents the range of 200 to 270 million passengers per month.
4.2 Data Series Analysis

In this study, we analyze the time series formed by the quantity of passengers transported by the urban bus system in the São Paulo City from January 2005 to December 2023. The data is divided into two periods: the first spans from January 2005 to February 2020 (pre-pandemic), and the second covers March 2020 to December 2023 (post-pandemic).

For the time series analysis and forecasting, we employ the ARIMA (AutoRegressive Integrated Moving Average) method using the Jamovi software, based on the R programming language. As described by Hyndman and Athanasopoulos (2018), ARIMA involves correlating current and lagged values of a data series with lagged values of random error terms. If necessary, differencing is applied to address non-stationarity. The ARIMA model is defined by three parameters: p (order of autoregressive component), d (degree of differencing), and q (order of moving average component). When dealing with seasonal time series, the SARIMA (Seasonal ARIMA) model combines non-seasonal (p, d, q) and seasonal (P, D, Q) components, denoted as SARIMA(p,d,q)(P,D,Q)s or ARIMA(p,d,q)(P,D,Q)s, where s represents the seasonal period.

Our methodology comprises three steps: model identification, parameter estimation, and model adequacy verification. We adopt an automated approach, allowing the software to select the best models and configurations while focusing on assessing their fit to the data reality.

5 RESULTS AND DISCUSSIONS

The initial step of our study involved analyzing the decomposition of the time series, as illustrated in Figure 6. This decomposition allowed us to discern the observed series itself, trend components, seasonal patterns, and random fluctuations within the data. Notably, the series exhibited an initial upward trend, which subsequently reversed into a declining trend. Additionally, clear seasonality was evident at the beginning and end of each period (January, February, and December), as well as in the middle of the period (July)—corresponding to school vacations in São Paulo.

Subsequently, we assessed the conditions of stationarity and autocorrelation using graphical analysis of the autocorrelation function (ACF) and partial autocorrelation function (PACF). The ACF revealed non-stationarity (Figure 7[a]), with significant autocorrelations at various lags. Meanwhile, the PACF (Figure 7[b]) exhibited strong correlations at lags 1, 5, and 12, along with weaker correlations at other lags. Recognizing the need for data differencing, we requested the software to determine the necessary order of differencing, resulting in a single differentiation step. Figure 8 displays both the original series and the series after applying this differentiation.

Following differentiation, we re-evaluated the ACF (Figure 9[a]) and PACF (Figure 9[b]) of the transformed series, which now demonstrated stationarity. Subsequently, we conducted the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests for stationarity. The ADF test requires a p-value < 0.05, while the KPSS test necessitates a p-value > 0.05 to confirm data stationarity, thereby rejecting the null hypothesis of non-stationarity.
Figure 6. Decomposition of the Monthly Commute Time Series

Source: Compiled by the Authors.

Figure 7. Autocorrelation Functions (ACF) and Partial Autocorrelation Functions (PACF) Before Differencing.

Source: Compiled by the Authors.

Figure 8. Original Data Series and Post-Differencing Series.

Source: Compiled by the Authors.
As observed in Table 1, initially, both tests indicated non-stationarity in the data series. However, after applying data differencing, both tests confirmed the stationarity of the dataset.

Table 1. Results of ADF and KPSS Stationarity Tests

<table>
<thead>
<tr>
<th>Differentiation</th>
<th>Test</th>
<th>Statistics</th>
<th>Lag*</th>
<th>p-value</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without</td>
<td>DFA</td>
<td>-2,8114</td>
<td>5</td>
<td>0,237</td>
<td>p-value &gt; 0,05</td>
</tr>
<tr>
<td></td>
<td>KPSS</td>
<td>0,8137</td>
<td>4</td>
<td>&lt; 0,01</td>
<td>p-value &lt; 0,05</td>
</tr>
<tr>
<td>With</td>
<td>DFA</td>
<td>-9,1891</td>
<td>5</td>
<td>&lt; 0,01</td>
<td>p-value &lt; 0,05</td>
</tr>
<tr>
<td></td>
<td>KPSS</td>
<td>0,2338</td>
<td>4</td>
<td>&gt; 0,10</td>
<td>p-value &gt; 0,05</td>
</tr>
</tbody>
</table>

* For DFA Lag = Lag Order; for KPSS Lag = Truncation Lag Parameter

In accordance with these definitions, the software generated an ARIMA(5,1,0)(2,1,2)[12] model. Specifically, this model comprises 5 autoregressive terms (p) for the non-seasonal component, 2 seasonal autoregressive terms (P), 1 first-order differencing term (d) for the non-seasonal component, 1 seasonal differencing term (D), no simple moving average terms (q), 2 seasonal moving average terms (Q), and a seasonal period (s) of 12 months. The resulting model yielded an AIC (Akaike Information Criteria) value of 5,830.0159 and a BIC (Bayesian Information Criteria) value of 5,861.3149.

The final verification step involved analyzing the residuals (Figure 10), which exhibited a well-distributed pattern around zero. This observation was further supported by the Shapiro-Wilk test (p-value < 0.05). The model’s forecast (Figure 10) closely aligns with the pre-pandemic data series (Figure 3). In detail, Figure 11 displays the post-pandemic period, including a 95% confidence interval (orange lines). Notably, during 2020 and 2021, the projected values (red line) significantly exceeded the observed data (gray line), a phenomenon readily explained by pandemic-related restrictions. However, in 2022 and 2023, there is a remarkable similarity between the projected and observed variations, with the observed data gradually approaching the model’s predictions.
The discrepancy between projected and observed passenger demand warrants investigation, particularly in the context of the COVID-19 pandemic. The monthly average forecast for 2023 hovers around 206 million passengers, while the observed monthly average remains at approximately 173 million passengers—a 19% shortfall. Comparing this to 2019, when the monthly average reached around 220 million passengers, the 2023 demand falls 21% below historical levels.

Examining the total annual passengers transported in the post-pandemic period (Table 2), we observe a substantial reduction in the discrepancy between projected and observed demand from 2020 to 2023. However, there remains a significant gap to bridge to reach the projected 2023 value. Notably, this projected value is relatively low when compared to the historical series dating back to 2005 (as depicted in Figure 13).

<table>
<thead>
<tr>
<th>Year</th>
<th>Demand Observed (Millions)</th>
<th>Demand Forecasted (Millions)</th>
<th>Difference Absolute (%)</th>
<th>Annual variation Observed</th>
<th>Annual variation Forecasted</th>
<th>Base 2011 variation observed</th>
<th>Base 2011 variation Forecasted</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019</td>
<td>2,638,190,764</td>
<td>2,573,022,432*</td>
<td>-5.7%</td>
<td>-10.3%</td>
<td>-10.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td>1,563,252,681</td>
<td>1,009,769,751</td>
<td>-40.7%</td>
<td>-46.8%</td>
<td>-12.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2021</td>
<td>1,674,525,550</td>
<td>879,158,603</td>
<td>-34.4%</td>
<td>-43.1%</td>
<td>-13.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td>2,050,203,520</td>
<td>458,610,668</td>
<td>18.3%</td>
<td>-30.3%</td>
<td>-14.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2023</td>
<td>2,079,105,034</td>
<td>396,821,403</td>
<td>16.0%</td>
<td>-29.3%</td>
<td>-15.8%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The forecasted demand for 2020 includes the values of the demand observed in January and February

Source: Compiled by the Authors.
In the total annual passengers transported during the post-pandemic period (Table 3), we observe a reduction in the discrepancy between projected and observed demand in 2023. However, there remains a considerable gap to bridge to reach the projected 2023 value, which is relatively low when compared to the historical series (15.8% below the 2011 demand). Figure 12 illustrates that the projected demand (shaded lines) exhibited a continuous decline, albeit less pronounced than the observed annual decline in 2019 (5.7%). The average annual projected decline from 2011 to 2023 would be 15.8%, corresponding to a loss of over 35 million passengers or trips per year in the São Paulo City bus system—from the peak value in 2011 to the forecast for 2023. Comparing the same period (2005 to 2023), the decline would be only 1.2%, equivalent to an average annual loss of 2.4 million passengers, with 2023 representing the lowest value in the annual historical series.

Figure 12. Annual Observed and projected Demand Trends (Solid Lines and Shaded Lines).

5 CONCLUSIONS

The detailed analysis of distinct phases of the COVID-19 pandemic and their impact on urban mobility underscores the urgent need to rethink policies and practices related to public transportation. The decentralized approach taken by states and municipalities during the pandemic revealed the complexity of managing mobility in a global crisis context. Notably, the three waves of COVID-19 in Brazil exacerbated the challenges faced by urban public transportation, which was already experiencing declining demand even before the pandemic.

The shift in mobility patterns during the pandemic—favoring individual modes and drastically reducing public transportation demand—poses an additional challenge to system sustainability. Consequently, beyond post-COVID economic recovery, redefining strategies to make public transportation more attractive and aligned with users’ evolving behavioral dynamics becomes crucial, as indicated by post-pandemic analyses.

Examining passenger data from São Paulo’s municipal bus system, which has undergone various transformations over the years (including municipalization in 1991, the introduction of the BU in 2004, and integration with rail transport in 2005), reveals a continuous decline in passenger demand with significant year-to-year variations. The data clearly reflects
the abrupt decline at the pandemic’s onset, followed by gradual recovery as the pandemic unfolded and eventually subsided. However, observed demand still lags behind projections.

Monthly and seasonal analyses highlight consistent behavioral patterns up to 2019, with peaks in March, May, August, and October. The arrival of the Covid-19 pandemic in 2020 led to a sharp drop in demand, evident in the time series analysis. We applied an ARIMA(5,1,0)(2,1,2)[12] model to predict demand behavior through the end of 2023, assuming the pandemic had not occurred.

Comparing projected and observed demand reveals a significant discrepancy attributable to ongoing pandemic impacts. Although the projected decline trend is less pronounced than the annual drop observed in 2019, the system has not yet reached projected 2023 levels. Annual analysis shows a substantial average passenger loss, posing a challenge for São Paulo City’s recovery.

This study underscores the significant challenges faced by passenger demand in São Paulo’s municipal bus system, influenced by factors such as the pandemic and potential shifts in population mobility habits. Implementing strategies to revitalize the system and attract passengers is critical for ensuring the city’s public transportation remains sustainable and adaptable to social and economic transformations.

BIBLIOGRAPHY


