Journal of Infrastructure, Policy and Development 2024, 8(11), 7918.
https://doi.org/10.24294/jipd.v8i11.7918

Article

Linda index forecast based on ARIMA time series

Cesar Hernández*, **Jesús Prieto, Diego Giral**

Faculty of Technology, Universidad Distrital Francisco José de Caldas, Bogotá 111611, Colombia
*** Corresponding author:** Cesar Hernández, cahernandezs@udistrital.edu.co

CITATION

Hernández C, Prieto J, Giral D. (2024). Linda index forecast based on ARIMA time series. Journal of Infrastructure, Policy and Development. 8(11): 7918. https://doi.org/10.24294/jipd.v8i11.7918

ARTICLE INFO

Received: 18 September 2024
Accepted: 27 September 2024
Available online: 16 October 2024

Abstract: Competition in the telecommunications market has significant benefits and impacts in various fields of society such as education, health and the economy. Therefore, it is key not only to monitor the behavior of the concentration of the telecommunications market but also to forecast it to guarantee an adequate level of competition. This work aims to forecast the Linda index of the telecommunications market based on an ARIMA time series model. To achieve this, we obtain data on traffic, revenue, and access from companies in the telecommunications market over a decade and use them to construct the Linda index. The Linda index allows us to measure the possible existence of oligopoly and the inequality between different market shares. The data is modeled through an ARIMA time series to finally predict the future values of the Linda index. The results show that the Colombian telecommunications market has a slight concentration that can affect the level of competition.

Keywords: ARIMA; concentration; Linda index; market; modeling; time series; telecommunications

1. Introduction

The services of the telecommunications market have a strong impact on important sectors of society such as education with access to the internet and information and communication technologies, health with the strengthening of telemedicine, and the economy through technologies that increase productivity, among others (Aguilar et al., 2020; Bardey et al., 2022; Bravo et al., 2022; Comisión de Regulación de las Comunicaciones, 2023a). The development of the telecommunications industry has been experiencing strong growth driven by technological innovation and strengthened by market demand and government policies (Grand View Research, 2021; Wang and Liu, 2022).

Studies have been carried out to analyze the current status of the telecommunications market in Latin American countries and to identify the need for elements such as adequate spectrum management and alignment with established policies on telecommunications services. As a result of these studies, the analysis of competition in these markets has been identified as a central point of study to improve competitiveness and reduce the digital divide, facilitating regional development and identifying potential investments. It is determined that countries that allocate greater bandwidth and achieve more competitive market structures obtain a higher number of demonstrable social benefits. Countries such as Mexico have initiated this area, and the analysis of the spectrum, its allocation, and management is included among the essential variables as a relevant topic of study, appropriating the lessons of more mature markets (Bravo et al., 2022; Comisión de Regulación de las Comunicaciones, 2022; de Miera Berglind, 2015).

The above highlights the importance of adequate regulation in the telecommunications market to ensure competition in the telecommunications sector companies to positively impact the productivity of companies in all sectors and thus lead to the country's socioeconomic development. However, there is still much to do in countries like Colombia. All of the above is the main motivation for the development of this article. As mentioned above, the concentration of spectrum by a telecommunications operator leads to less competition in the market. It increases the digital divide, affecting the development of the country and the region. With this in mind, this work's added value and main objective is to measure the current concentration level in the Colombian telecommunications market. This allows a competition analysis to propose effective strategies and methodologies to improve competition in the Colombian telecommunications services operators (Comisión de Regulación de Comunicaciones, 2021; MinTIC, 2023).

However, predicting its behavior would be even better, since this would allow planning in advance the necessary strategies to regulate this market (Li et al., 2022; MinTIC, 2023; Nekmahmud, 2018). This paper aims to analyze concentration prediction in the telecommunications market. To achieve this, we worked with real data from companies in the Colombian telecommunications market; between 2012 and 2022, the Linda index was used to measure the level of concentration of companies in the telecommunications sector, and the prediction of this index was made through ARIMA time series (Mills, 2019). One of the most notable contributions of this work is the prediction of the Linda index as a method of anticipation, forecasting and early planning for regulating telecommunications markets.

This work is organized and presented in five sections. Section 2 presents the research materials and methods. Section 3 describes the results obtained. Section 4 discusses the results. The final section, section 5, establishes the general conclusions of the work.

2. Materials and methods

The methodology is described based on the three main activities to develop the proposed model: telecommunications market data; the Linda market concentration index; and the ARIMA time series model.

2.1. Telecom market data

Initially, the data corresponding to the analysis variables were obtained, such as: traffic, revenues and accesses, corresponding to each of the telecommunications companies that operated and operated in Colombia from 2012 to September 2022 (inclusive); this information was obtained from the Communications Regulatory Commission's postscript database. Subsequently, the data was organized in Excel to create a database with the information of interest organized chronologically. Here, nine databases were finally obtained: (1) traffic from fixed-charge (postpaid) mobile internet demand; (2) revenue from fixed-charge (postpaid) mobile internet demand; (3) accesses of the fixed-charge (postpaid) mobile internet demand; (4) mobile internet demand traffic on demand on demand (prepaid); (5) revenue from mobile internet demand on demand (prepaid); (6) mobile internet demand accesses on demand (prepaid); (7) global mobile internet demand traffic (postpaid + prepaid); (8) revenue from global mobile internet demand (postpaid + prepaid); (9) accesses of global mobile internet demand (postpaid + prepaid). For this work, databases (7), (8) and (9) were used, to handle global information (postpaid + prepaid) (Comisión de Regulación de Comunicaciones, 2023b).

Figure 1 describes the behavior of traffic data for Comcel, one of the most important companies in Colombia in the telecommunications market. The behavior of the data for the variables revenue and access can be seen in **Figures 2** and **3**, respectively.

Subsequently, the traffic, entry and access data of each company in the country's telecommunications sector were organized to calculate the value of the Linda concentration index.

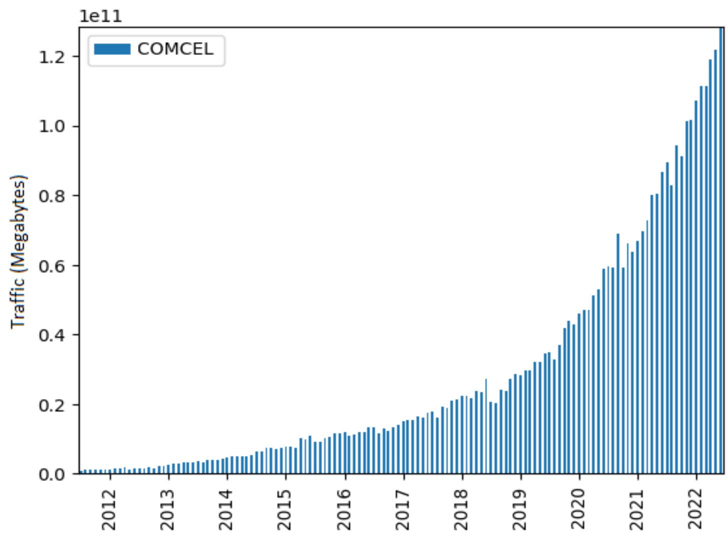

Figure 1. Traffic data for the company Comcel.

Source: Authors.

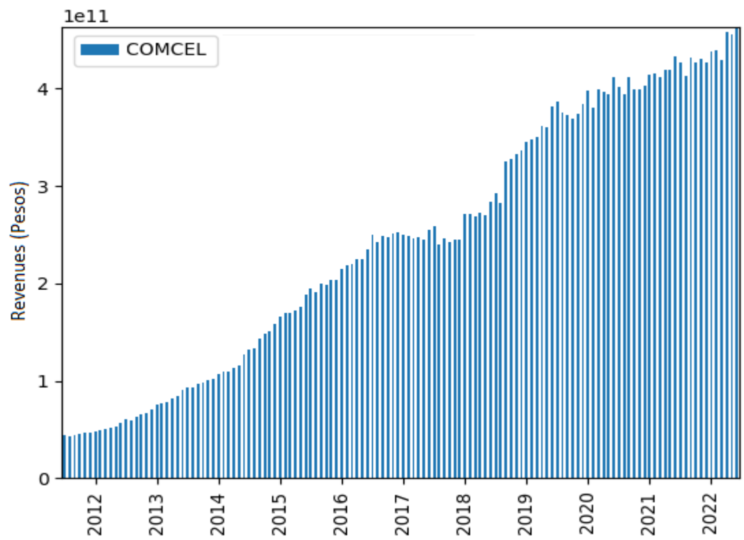

Figure 2. Revenue data for the company Comcel.

Source: Authors.

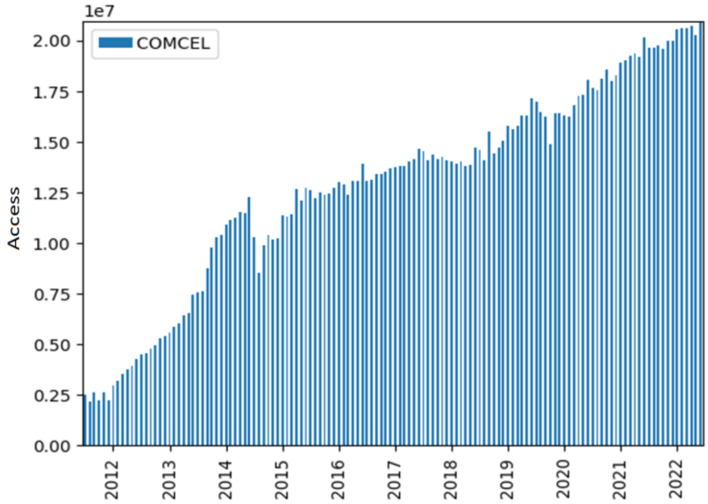

Figure 3. Access data for the company Comcel.

Source: Authors.

2.2. Linda index

This indicator is usually used to measure the possible existence of oligopoly and inequality between different market shares. In addition, similar to the concentration ratio, it is calculated for a number n of leading companies in the market, so that their joint relative incidence about the rest of the participants at that end of the market (supply or demand) can be calculated (Comisión de Regulación de Comunicaciones, 2023; Lis-Gutiérrez, 2013).

This index focuses on the distribution of the k companies with the largest market share; since it is designed to evaluate the degree of inequity in a market and the presence of oligopolies. So, unlike the previous indexes, it compares the concentration between two groups of companies, those with the largest market share (leaders) and the others, allowing one to calculate their joint relative incidence about the rest of the companies (Apolinario et al., 2022). This indicator can be defined mathematically as shown in Equation (1).

$$L = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \frac{\bar{X}_{\iota}}{\bar{X}_{N-i}} \tag{1}$$

where:

$\bar{X}_{\iota}$ is the average market share of the first *i* companies.

$\bar{X}_{N-i}$ is the average market share of the remaining companies.

N is the total number of companies in the market.

The range of values of Linda's index is between zero and infinity, and its interpretation can be seen in **Table 1** (Melnik et al., 2008).

This indicator presents values between zero and infinity, with values close to zero obtained by markets with low concentration and values close to one (especially greater than one) representing highly concentrated markets, as shown in **Table 1** (Comisión de Regulación de Comunicaciones, 2023b).

Table 1. Interpretation of Linda's index.

Concentration	Rank
Casualty	<0.2
Moderate	0.2 to 0.5
Loud	0.5 to 1
Very High	>1

Source: Authors.

Since forecasts of the Linda index are planned later, it was decided to calculate this index every month to obtain a greater amount of data. The procedure required to calculate the Linda index, (1) it was necessary that for each period (month), the telecommunications companies were ordered from highest to lowest according to the value of the variable to be analyzed (traffic, revenues or accesses); (2) the percentage of participation of each company is calculated for each period, for each variable analyzed, so that the value of the sum of the participations for each period is 1; (3) for each period, two groups of companies are formed, the first group formed by the two companies with the largest participations, and the second group with the rest of the companies; (4) for each group the average participation of said group is calculated; (5) the result of the average participation of the first group is divided by the result of the second group, and multiplied by $1/(N \times (N - 1))$, where N is the number of companies in each period.

Finally, the Linda index was calculated for each database mentioned above. Here it was evident that the Linda index was indeterminate for periods where some companies had zero values in the variable of interest. If it was very close to zero, it increased exponentially. Due to the above, it was determined to eliminate the data equal to zero, since the interpretation of these is fundamentally, that the company did not operate in that period. Additionally, it was decided to eliminate all data less than 50,000 in the traffic and revenue databases, both postpaid and prepaid and global, the amount of data deleted was 36, which gives approximately a value of less than 0.4% of the total database.

For the months in which more than three companies were operating, more than one Linda index was obtained, since this index compares groups of companies, the first Linda that is obtained is given if at least three companies are competing in the market and would correspond to Linda 2 (L2), if there are four companies L2 and L3 would be obtained, and so on, i.e. the last Linda corresponds to $N - 2$, where N is the number of companies competing in the market. Because there are periods where up to 14 companies operate simultaneously, Linda indices of up to L12 are obtained. Graphing each would represent an extension of this document, so it was decided to graph only the Linda 2 (L2) of each period.

Figures 4–6 present the Linda index of global mobile internet demand for traffic, revenue and access, respectively.

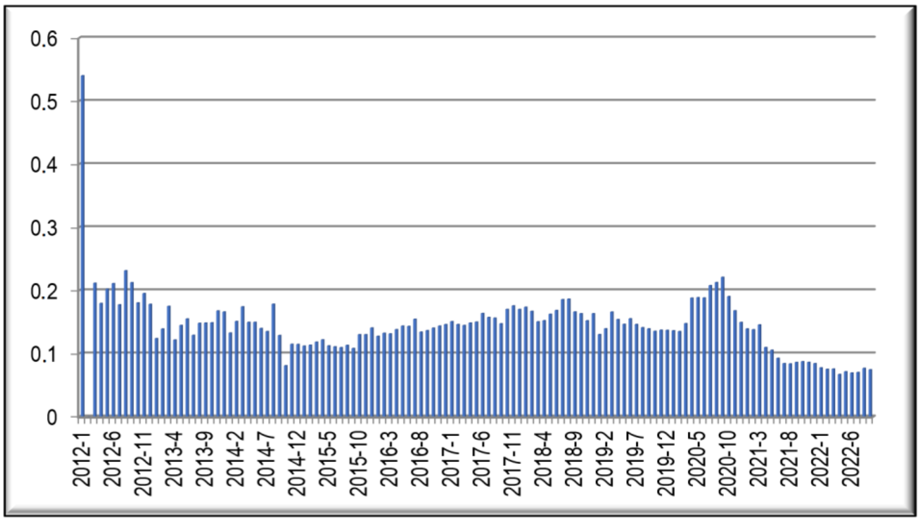

Figure 4. Linda index values for traffic of global mobile internet demand.

Source: Authors.

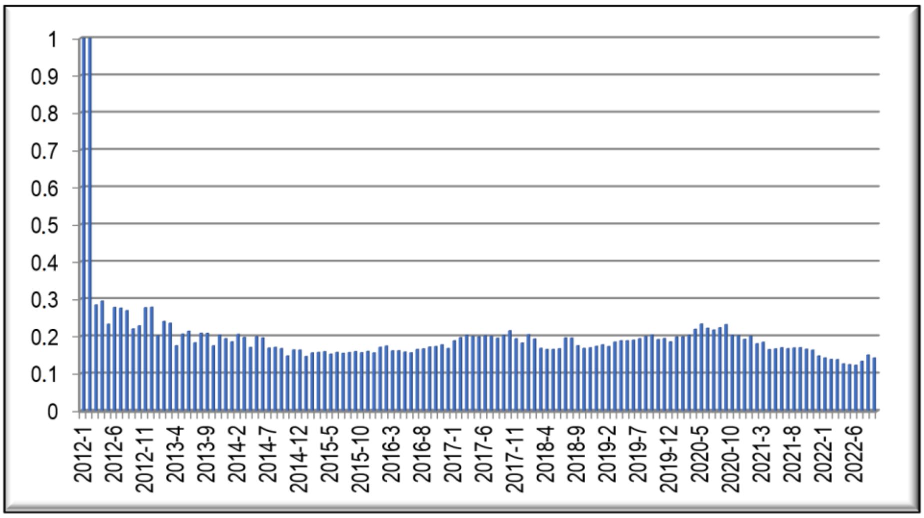

Figure 5. Linda index values for revenue of global mobile internet demand.

Source: Authors.

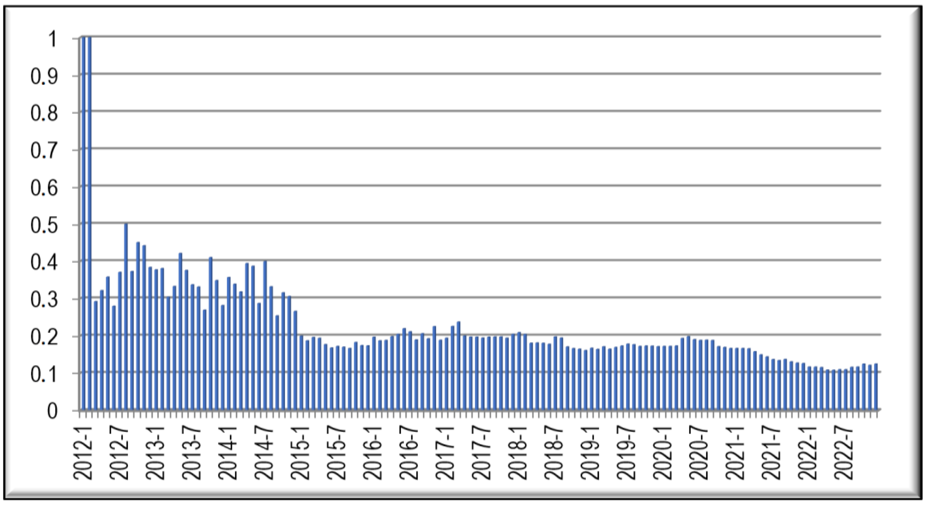

Figure 6. Linda index values for the number of accesses of global mobile internet demand.

Source: Authors.

2.3. Time series

A time series is a sequence *Xt*, generated by obtaining one and only one observation of each of the random variables that define a stochastic process. The observations are taken at equal time or distance intervals, as indicated by the index *t* that generates the sequence. In this sense, the series is a realization of a stochastic process (Yasmin, 2024).

2.3.1. Stationarity

Formally, a variable *Xt* is stationary if the mathematical expectation of the variable *Xt* (*E* [*Xt*]) is a constant, for all values of *t*.

The variance of the variable *Xt* (VAR [*Xt*]) is a constant, for all values of *t*.

The covariance of the product *Xt* × *Xt* + *k* (Cov [*Xt* × *Xt* + *k*]) is a constant for all values of *t* and all *k*, respectively.

These conditions require that the means, variances and covariances of *Xt* remain constant over time, meaning that it does not matter if the observations come from the beginning or the end of the sample, as long as the means and variances are always the same.

In other words, *Xt* will be stationary only if it is normally distributed with mean 0 and constant variance, that is: *N* (0, S2) (Ray et al., 2023).

2.3.2. Seasonality

Seasonality represents a periodic movement of the time series. The length of the period unit is generally less than a year. It can be a quarter, a month or a day, etc. (Zhang et al., 2024).

Mathematically, we can say that the time series has a seasonal component if there exists a number "s" such that *X*(*t*) = *X*(*t* + *ks*).

2.3.3. Forecast

Prediction of a value of a random variable, $X_{n+\ell}$, of a stochastic process.

n is the origin of the forecast and ℓ the relative position of the variable to be forecast with respect to *n*.

Generally, one is interested in forecasts of future random variables, that is, in predicting values that *Xt* takes when *t* > *n* (or when $\ell > 0$, if n is considered as the origin or as the present value) (H. Wang et al., 2023).

In the time of Galileo, Ptolemy or Copernicus, they were either impossible to predict or had very large margins of error.

It is concluded, then, that a forecast system, in addition to predicting a future value of a random variable, must estimate its error and ideally its distribution, which will allow the risk in decision making to be quantified (X. Wang et al., 2023).

2.3.4. Residual analysis

The differences between the observed and fitted data are known as residuals. Once a particular model has been fitted to a given time series, the residuals can be plotted over the n periods. If the particular model fits adequately, the residuals represent the irregular component of the time series and should therefore be randomly distributed throughout the series. On the other hand, if the particular model does not fit adequately, the residuals may be pointing to some systematic pattern such as a failure to explain the trend, a failure to explain cyclical variation, or, with monthly data, a failure to explain seasonal variation (Kumari and Muthulakshmi, 2024).

2.4. ARIMA model

In a forecast, there is always a latent uncertainty caused by the future, which is reflected in an increase in the prediction error. Therefore, it is important to achieve a minimum reduction in the prediction error. Time series stand out among other prediction methods for their low level of computation and their high level of precision in their forecasts. Within the time series models, the Auto-Regressive (AR), Moving Average (MA), Auto-Regressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA) models stand out. However, if the series to be forecasted is not stationary, the AR, MA and ARMA models would not work properly, as is the case with the present work, so the most appropriate models are ARIMA and SARIMA, which allow the series to be transformed to make it stationary and to be able to model it properly. In this work, the ARIMA model will be used to model the time series of interest and the SARIMA model will be left as future work.

The ARIMA time series was used to predict the Linda index corresponding to the variables of traffic, revenue, and accesses of the Comcel company (Box et al., 2016; Chávez, 1997). This time series does not present a high computational level, but it does have an excellent level of prediction, according to previous work (Akinrotimi et al., 2023; Alizadeh et al., 2023). This time series model was developed in Matlab (Mathworks, 2024).

The ARIMA model is derived from its three fundamental components: AR (Autoregressive), I (Integrated), and MA (Moving Averages). The ARIMA model describes the present value of a stochastic variable as a linear function of previous values and random errors. It can incorporate cyclical or seasonal components, that is, this model is conceived integrally, ensuring the inclusion of all the necessary elements for a complete description of the phenomenon studied (Fuentes; 2023). This approach allows us to understand not only the relationship between observations at specific times but also to capture possible cyclical or seasonal patterns that may influence the behavior of the data over time (Box et al., 2016; Brockwell and Davis, 2002).

It is important to note that the time series must be stationary to employ the ARMA model, as this ensures that the statistical properties of the time series will not change over time, bringing stability and reliability to parameter estimates and predictions. However, it should be noted that, in the case of having a non-stationary series, the ARIMA model must be used, which allows working with this type of series by including differentiation steps to make them stationary.

A series of organized steps are followed to obtain the ARIMA model of a time series. The first step is determining whether the time series you want to model is stationary. In case it is not, a series transformation is carried out to make it stationary, one of the most used transformations is differentiation, which can be performed one, two or more times until the series becomes stationary. The number of times the series is differentiated or transformed determines the 'd' value of the integrated component (I). Subsequently, the ARIMA model is identified, that is, the order of its autoregressive component (AR), called 'p' and its moving average component (MA), called 'q', is determined. Once the ARIMA model has been identified, the estimation of its coefficients is carried out. Finally, the model is validated through its residuals, that is, the values of difference between the real value and the value estimated by the model; When these residuals do not present any type of correlation, the model is said to be correctly estimated. If the above does not happen, the model is re-identified including the orders of p and q that were identified during the correlation test of the residuals.

2.4.1. Stationarity test

The Dickey-Fuller unit root test was used in the three-time series of traffic, revenues and accesses to determine whether the time series are stationary (see **Table 2**). All the tests showed that the series was not stationary, so the differentiation was done to transform it into stationary.

Table 2. Dickey-Fuller unit root test.

Time Serie	Traffic	Revenue	Accesses
Dickey-Fuller unit root test	1.49785	1.92237	1.6274
Critical value of 5%	−2.867	−2.621	−2.754

Source: Authors.

Since the time series are not stationary, they are incompatible with the ARMA model, however, when opting for the ARIMA model, the I(d) component is automatically responsible for transforming them into stationary, depending on the successive d-differentiations.

2.4.2. Transformation

According to Equation (2), differentiation was necessary to transform the time series of traffic, revenues, and accesses.

$$\triangle^n X_t = \triangle^{n-1} X_t - \triangle^n X_{t-1} \tag{2}$$

where $\triangle^n X_t$ is the differential of order n at time t, $\triangle^{n-1} X_t$ is the value of the series at time t, and $\triangle^n X_{t-1}$ is the value of the series at time $t-1$. Finally, the variable $\triangle^n X_t$ represents the differentiation n at time t.

Three differences were necessary for the traffic variable until the series was stationary. The **Figure 7** shows the result of the third differentiation for the Comcel company. In the case of the variables revenue and access, it was only necessary to differentiate them once.

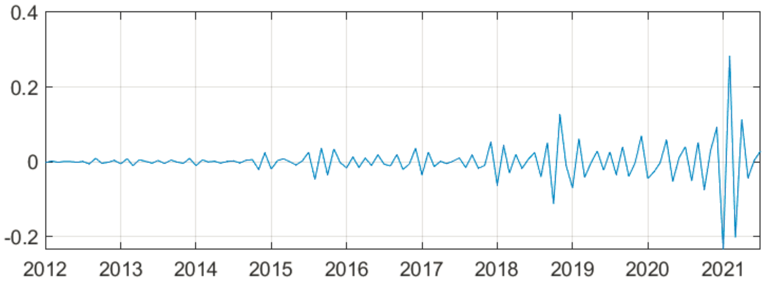

Figure 7. Third Traffic differentiation for the company Comcel.

In the case of the ARIMA model, identification consists of determining the value of the order of the AR, MA and I process. In the case of process I, these correspond to the number of times the signal was differentiated, three for traffic, and one for entrances and accesses. To find the value of p and q, the Partial Autocorrelation Function (PACF) and the Autocorrelation Function (ACF) are used. These functions play a crucial role in providing a visual and informative perspective on the temporal dependence of data.

The structure of an ARIMA model is presented in the Equation (3).

$$X_t^{(d)} = C + \varphi_1 X_{t-1}^{(d)} + \varphi_2 X_{t-2}^{(d)} + \cdots + \varphi_p X_{t-p}^{(d)} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

where (d) refers to differentiation, C to the constant, the blue part to the AR(p) model, the red part to the MA(q) model and refers to the white noise or error. In contrast, the error is defined as the disparity between our prediction and the correct value, also known as residuals, being possible to calculate it only after knowing the precise value of our variable of interest $\varepsilon_t X_t^{(d)}$.

Figures 8–10 present the results of the ACF function, for traffic, revenues and accesses, respectively.

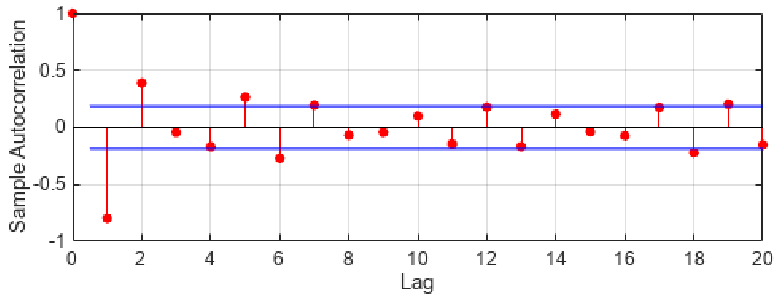

Figure 8. Autocorrelation function of traffic for the company Comcel.

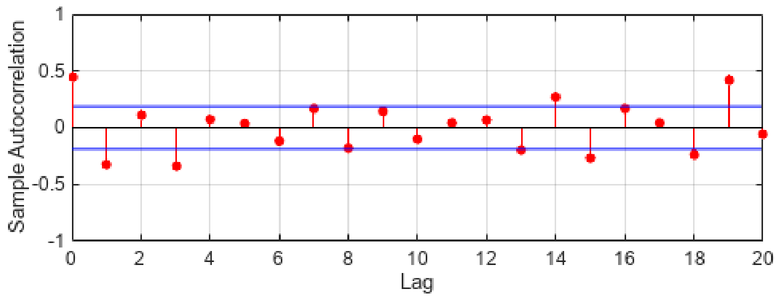

Figure 9. Revenue autocorrelation function for the company Comcel.

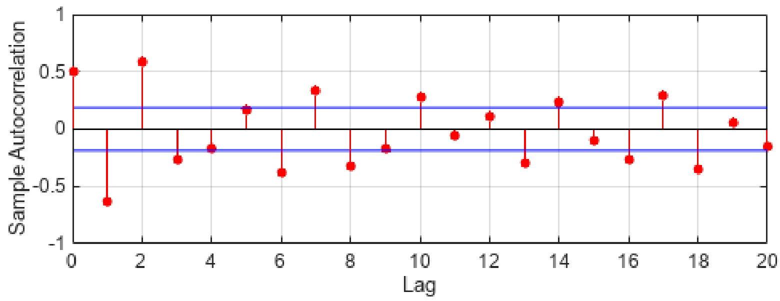

Figure 10. Access autocorrelation function for the company Comcel.

Figures 11–13 present the results of the ACFP function, for traffic, revenues and accesses, respectively.

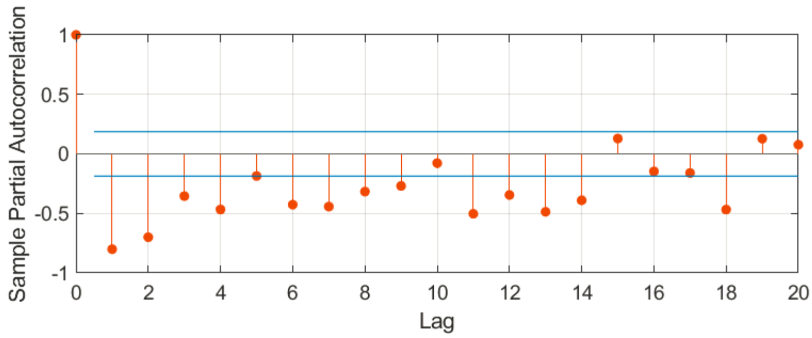

Figure 11. Partial autocorrelation function of traffic for the company Comcel.

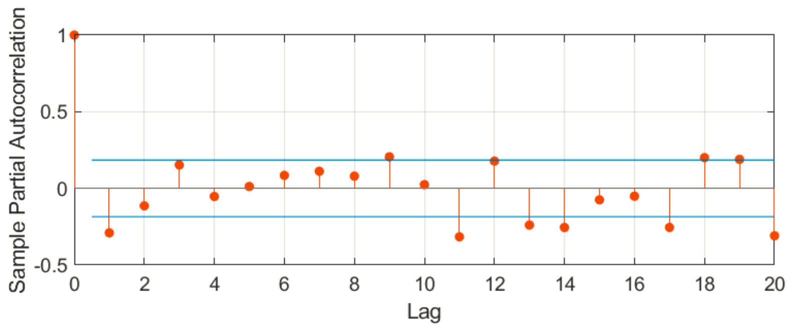

Figure 12. Partial revenue autocorrelation function for the company Comcel.

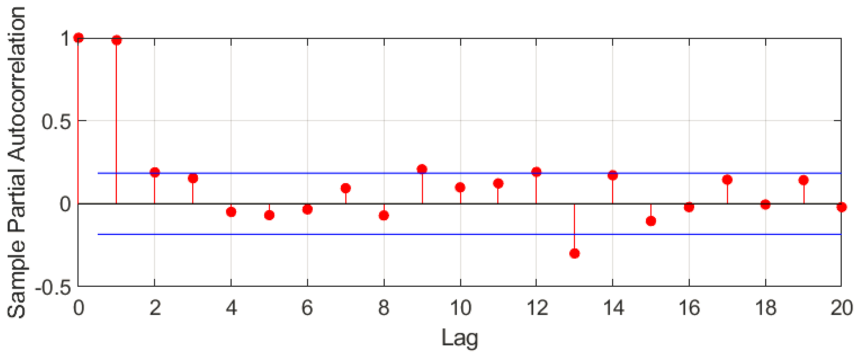

Figure 13. Partial autocorrelation of accesses function for the company Comcel.

The results of the ACF and PACF functions determine the values of p and q, i.e., the order of the AR and MA processes; here the lags that exceed the threshold shown in each of the ACF and PACF functions are taken into account, the furthest lag determines the order of p and q, respectively. After this, the estimation of the model's coefficients is carried out and validated through the correlation between its residuals. The process is iterative and is repeated until there is no correlation between the residuals. The models found at the end of this process are presented in Equations (4)–(6), for traffic, revenues and accesses, respectively.

$$X_t = 3X_{t-1} - 3X_{t-2} + X_{t-3} - 1.296 \times 10^{-5} - 1.967e_{t-1} - 0.986e_{t-2} + e_t \quad (4)$$

$$X_t = 0.008 - 0.016X_{t-1} + 0.777X_{t-2} + 0.520X_{t-3} + 0.194X_{t-4} - 0.183X_{t-5} - 0.434X_{t-6} + 0.141X_{t-7} - 0.758e_{t-1} - 0.051e_{t-2} + 0.171e_{t-3} + 0.443e_{t-4} - 0.441e_{t-6} + 0.022e_{t-7} + e_t \quad (5)$$

$$X_t = 0.886 - 0.057X_{t-1} + 0.776X_{t-2} + 0.262X_{t-3} + 0.086X_{t-4} - 0.076X_{t-5} - 0.886e_{t-1} + e_t \quad (6)$$

Table 3 presents the results of the Ljung-Box test, where it is evident that the *p*-values are greater than 0.05, so the null hypothesis is not rejected, demonstrating that the residuals of the models have no correlation, that is, they are independent. The above confirms that the developed models are adequate.

Table 3. Ljung-Box test.

	Traffic model	Revenue model	Accesses model
Ljung-Box test	6.45210	5.72532	6.00261
p-value	0.8423	0.6522	0.8048

3. Results

The developed models were validated and evaluated using 80% of the data for training and 20% for evaluating the predictions, which were unknown during the training phase.

90% of the data was used to validate and evaluate the developed ARIMA models, and the other 10% was used to evaluate the models' predictions, which were unknown during the training phase.

Figures 14–16 show the behavior of the ARIMA models developed concerning the data of the three-time series: traffic, revenue and access. **Figures 14–16** show that the blue line represents the time series and the estimates of the developed model are presented with an orange line. The model reflects a close coincidence between the two lines over time, which shows that the model has adequately captured the patterns and trends of the series. The modeling of the traffic time series shows an error of 1.63%, the revenue series of 1.64% and the access time series has an error of 2.73%.

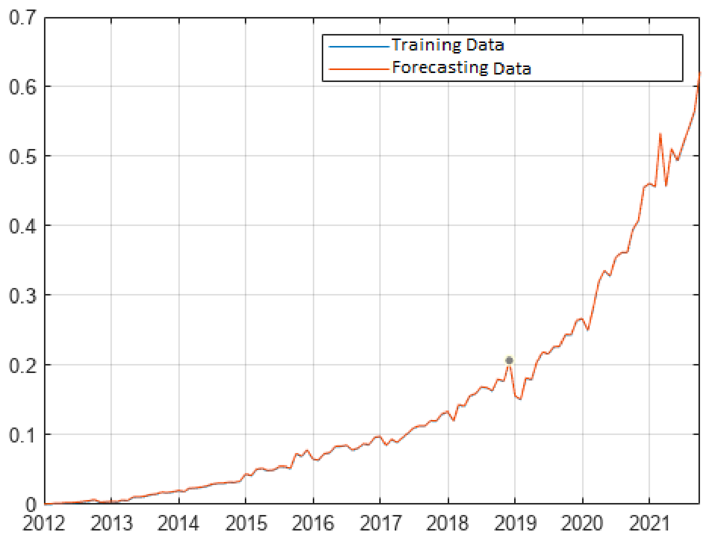

Figure 14. The behavior of the model for the traffic of the company Comcel.

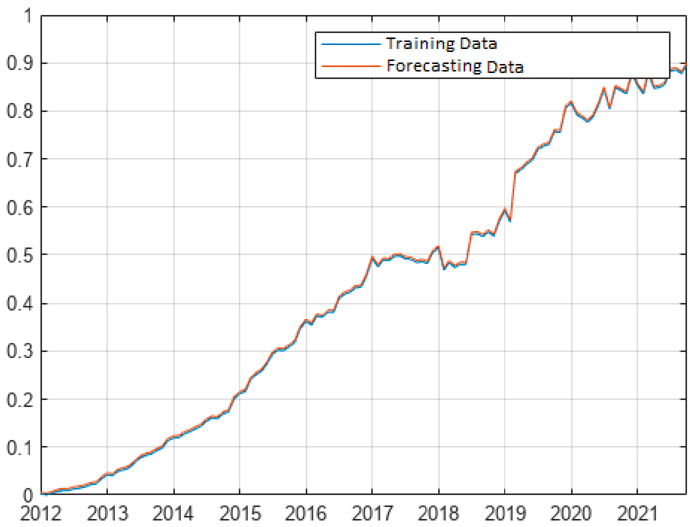

Figure 15. The behavior of the model for the revenues of the company Comcel.

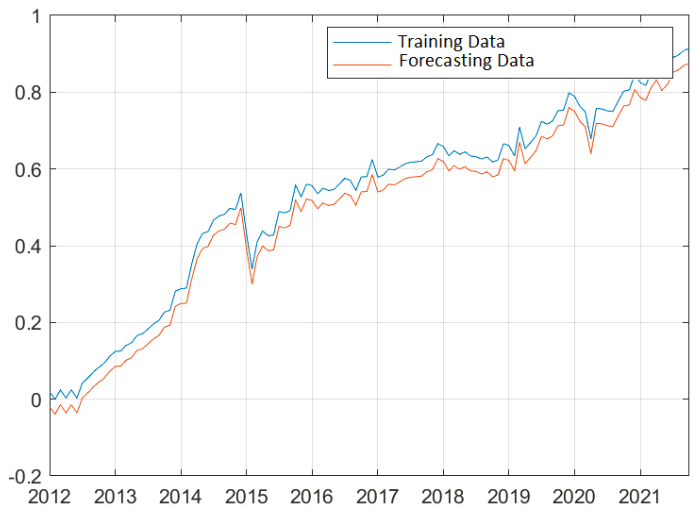

Figure 16. The behavior of the model for the accesses of the Comcel company.

For the prediction stage of the ARIMA models developed, 20% of the data was used to evaluate the accuracy percentage in the models' prediction.

Figure 17 describes the behavior of the ARIMA model's prediction for the Comcel company's traffic variable. The figure shows an excellent result since close values are observed when comparing the predicted and real data, and there is only a 2.77% error when evaluating the mean square error.

Figure 18 describes the behavior of the ARIMA model's prediction for the Comcel company's variable revenue. A value of 1.81% was obtained when evaluating the mean square error.

Figure 19 describes the behavior of the ARIMA model's prediction for the Comcel company's variable accesses. A value of 1.52% was obtained when evaluating the mean square error.

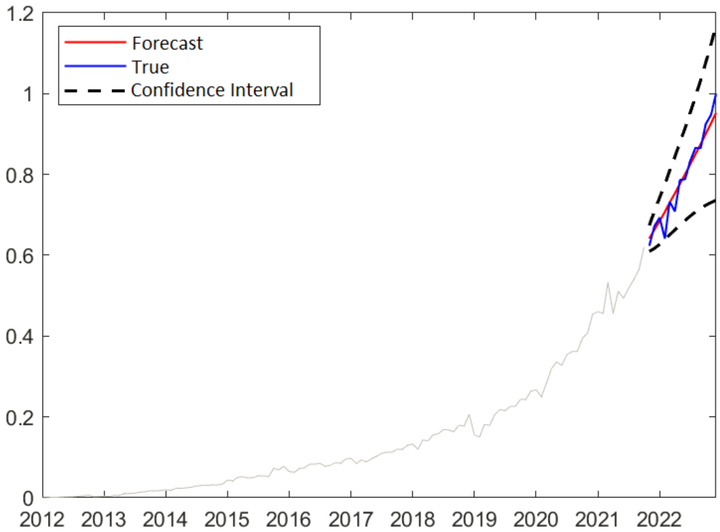

Figure 17. Traffic forecast for the company Comcel.

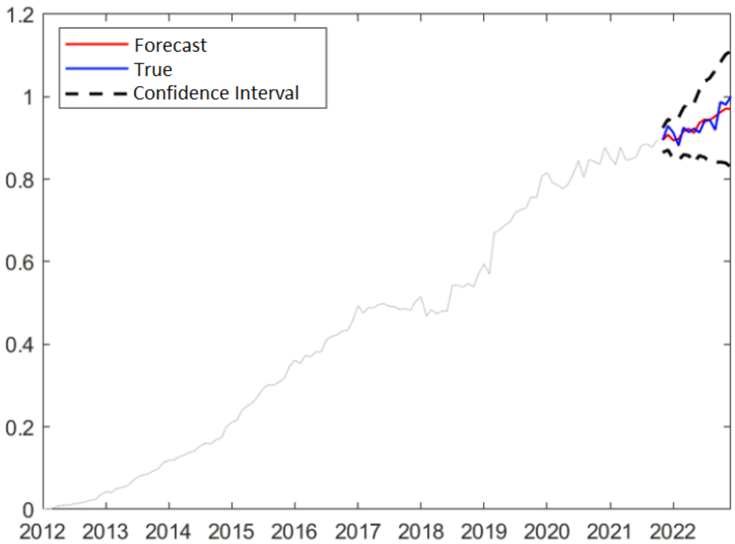

Figure 18. Revenue forecast for the company Comcel.

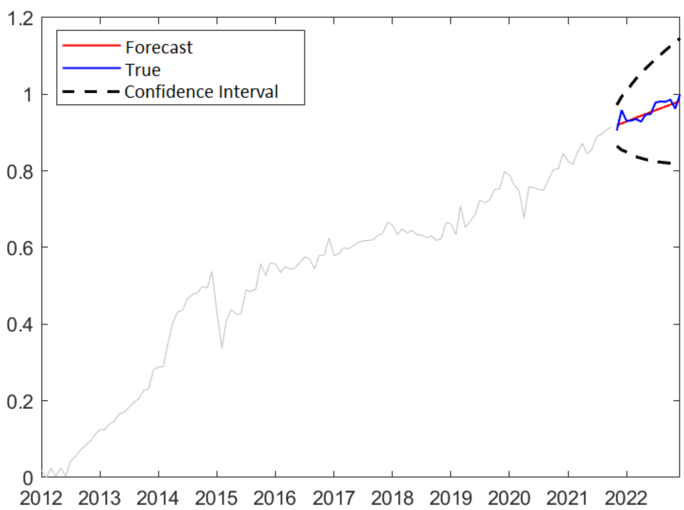

Figure 19. Access forecast for the company Comcel.

A process similar to that of the traffic, revenue and access time series was carried out, but now with the values of the Linda index for the same traffic, revenue and access variables. The results achieved are shown in **Figure 20**. According to the prediction, the index tends to increase for accesses, and after an increase it decreases for revenue and traffic.

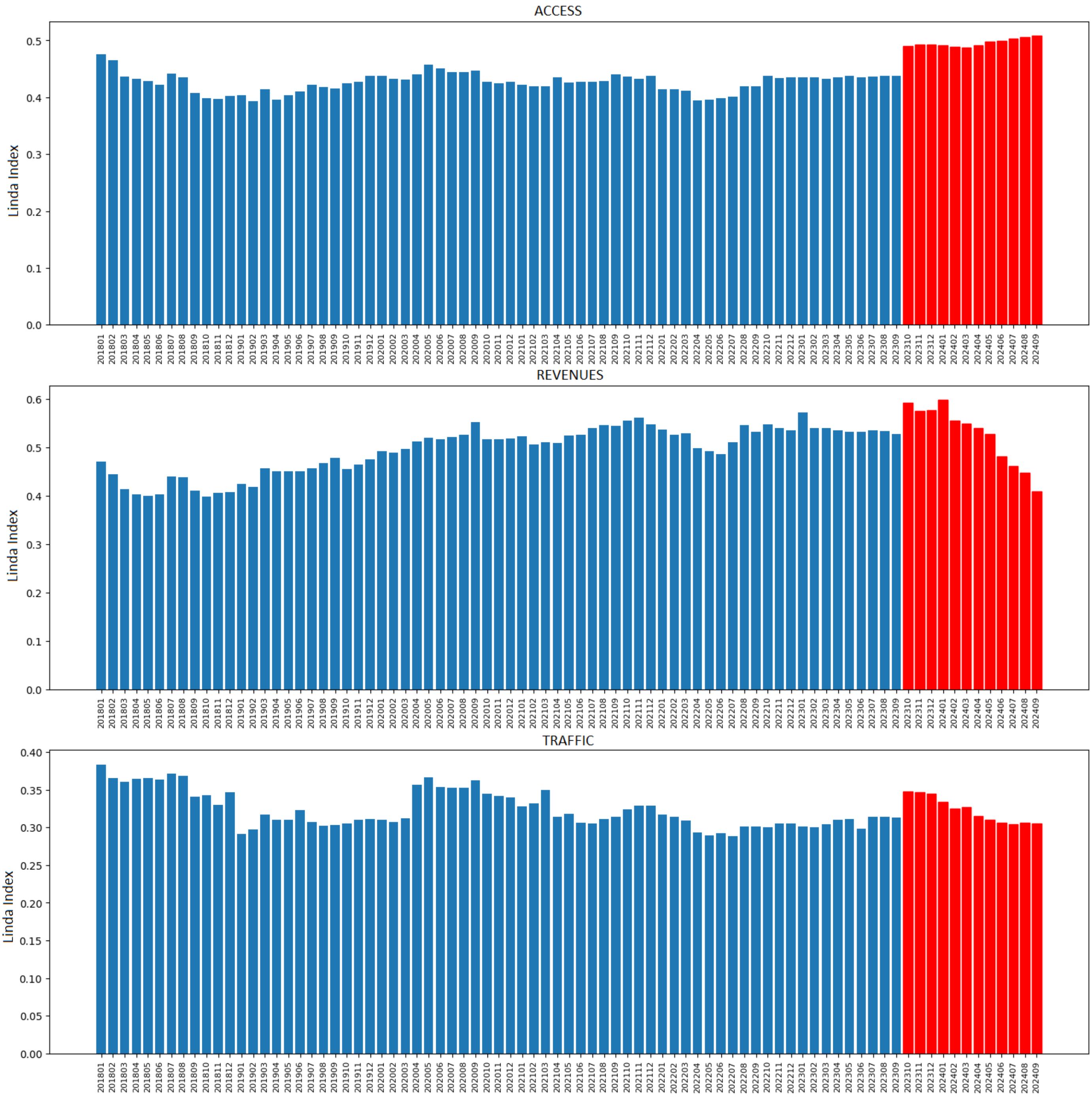

Figure 20. Linda index prediction for Comcel company traffic, revenues and accesses.

Source: Authors.

4. Discussion

The highest value of Linda 2 is given for the revenues of the global mobile internet demand (postpaid + prepaid) with a value of 21.11 and a standard deviation of 1.85, and for the number of accesses of the global mobile internet demand (postpaid + prepaid) with a value of 20.72 and a standard deviation of 2.49. The lowest Linda value is for global traffic with a value of 0.067 and a standard deviation of 0.0499, and for prepaid traffic with a value of 0.073 and a standard deviation of 0.111.

For the demand for fixed charge (postpaid) mobile internet, the values of the Linda index on average range between 0.23 and 0.32, indicating a moderate concentration. In the case of mobile internet demand on-demand (prepaid), the values of the Linda index on average vary, for traffic 0.15 (low concentration), for revenues 0.26 (moderate concentration), and for the number of accesses 0.41 (moderate concentration). In the case of global mobile internet demand, the values of the Linda index on average have greater differences: for traffic 0.14 (low concentration), for revenues 0.36 (moderate concentration), and the number of accesses 0.53 (high concentration).

It is important to emphasize that the Linda 2 index compares the groups of the two companies with the highest value of the variable of interest (traffic, revenues, or accesses) to those of other companies. It is possible that in some cases, there is evidence of a greater concentration for the group of the three or four most dominant companies in the market.

The division of the data for training and prediction allowed the effectiveness of the models to be evaluated. The low error rate in the traffic and revenue variables supports the appropriate and assertive selection of the model. At the same time, the discrepancy in the access data suggests possible areas for improvement in the adaptation of the model. This significant divergence could suggest the need to review and adjust this model. Ultimately, this analysis reveals that to effectively address the variations and complexities in access, revenue, and traffic data, it is crucial to tailor ARIMA models to the specific characteristics of each dataset. The resulting models accurately represent the temporal dynamics of each time series and provide a solid basis for decision-making and future planning in the context of the firm and by government regulation.

In conclusion, the proposed objectives were met, showing that the algorithm developed is a valuable tool for anticipating and understanding the dynamics of the Colombian market. The combination of careful technique selection, effective implementation in MATLAB, and rigorous evaluation of results provides a solid foundation for informed business and analytical decision-making. This allows the state to generate policies and strategies in advance that allow for the timely control of the emergence of oligopolies or monopolies, or mitigate them if they already exist to prevent their growth. This paper demonstrates the existence of a concentrated telecommunications market, where the level of competition is decreasing in the future, as shown by the projections of the Linda index.

5. Conclusion

The measurement and monitoring of competition in the telecommunications market allow the state to implement strategies to improve market regulation to avoid the formation of oligopolies and monopolies.

Linda’s traffic, revenue and access to mobile internet demand values show a significant concentration, indicating a need to improve competition in this market.

Time series are an excellent option for modeling temporal variables, especially ARIMA models that allow working with non-stationary series. Their level of complexity is not high, but their level of accuracy in forecasting is. The successful application of ARIMA models to forecast trends and patterns in traffic, revenue, and access validates the predictive power of the models. The visualization of results highlights the usefulness of these models to understand and anticipate future dynamics in the metrics analyzed.

This research demonstrates the existence of a concentrated telecommunications market, where the level of competition is decreasing in the future, as shown by the projections of the Linda index. This allows the state to generate policies and strategies in advance that allow for the timely control of the emergence of oligopolies or monopolies, or mitigate them if they already exist to prevent their growth.

The main limitation of this work was to be able to make comparisons with artificial intelligence models such as Deep Learning, which have recently gained great acceptance, and to propose an adaptive multi-model that would allow improving the results obtained in this research.

Future work aims to develop a hybrid model combining the results of at least three of the most widely used concentration indices to better predict concentration behavior in the telecommunications market. This will allow for the early proposal of effective strategies and methodologies to improve competition in Colombian telecommunications service operators.

Author contributions: Conceptualization, CH; formal analysis, methodology, writing—original draft, writing—review and editing, CH, JP and DG; project administration, CH; software, JP; validation, CH, JP and DG; data curation, DG. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministerio de Ciencia, Tecnología e Innovación (Minciencias) of Colombia and Universidad Distrital Francisco José de Caldas, with funding contract number 082-2022.

Acknowledgments: The authors thank the Investigations Office (ODI) of the Universidad Distrital Francisco Jose de Caldas and Minciencias for supporting this investigative project.

Conflict of interest: The authors declare no conflict of interest.

References

Aguilar, D., Agüero, A., & Barrantes, R. (2020). Network effects in mobile telecommunications markets: A comparative analysis of consumers’ preferences in five Latin American countries. Telecommunications Policy, 44(5), 101972. https://doi.org/10.1016/j.telpol.2020.101972

Akinrotimi, A. O., Ogundokun, R. O., Mabayoje, M. A., et al. (2023). A Smote-Based Churn Prediction System Using Machine Learning Techniques. 2023 International Conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG), 11, 1–6. https://doi.org/10.1109/seb-sdg57117.2023.10124631

Alizadeh, M., Beheshti, M. T. H., Ramezani, A., et al. (2023). An optimized hybrid methodology for short-term traffic forecasting in telecommunication networks. Transactions on Emerging Telecommunications Technologies, 34(12). Portico.

https://doi.org/10.1002/ett.4860
Bardey, D., Sáenz, B., Aristizábal, D., et al. (2020). Impact of the concentration of the mobile market in Colombia on Competitiveness (Spanish). Available online: https://economia.uniandes.edu.co/sites/default/files/eventos/Estudio-movistar.pdf (accessed on 12 August 2024).
Box, G. E. P., Jenkins, G. M., Reinsel, G. C., et al. (2016). Time Series Analysis: Forecasting and Control. John Wiley & Sons Inc.
Bravo Apolinario, Y., Quispe, D. K. P., Rodriguez, P. B. V., et al. (2022). Bank concentration, measured by different indicators. the Peruvian case. Journal of Globalization, Competitiveness and Governability, 16(1). https://doi.org/10.3232/gcg.2022.v16.n1.05
Brockwell, P. J., & Davis, R. A. (2002). Introduction to Time Series and Forecasting, Second Edition. Available online: http://home.iitj.ac.in/~parmod/document/introduction time series.pdf (accessed on 12 August 2024).
Chávez Quisbert, N. (1997). Arima Models. Revista Ciencia y Cultura, 1, 23-30.
Comisión de Regulación de Comunicaciones. (2023a). Postscript: Beyond the data (Spanish). Available online: https://postdata.gov.co/ (accessed on 12 August 2024).
Comisión de Regulación de Comunicaciones. (2023b). Battery of indicators for the analysis of competition in communications markets (Spanish). Available online: https://postdata.gov.co/story/bateria-de-indicadores-para-el-analisis-de-competencia (accessed on 12 August 2024).
Comisión de Regulación de las Comunicaciones. (2021). ICT and Postal Industry Report 2020 (Spanish). Available online: https://postdata.gov.co/ (accessed on 12 August 2024).
de Miera Berglind, O. S. (2015). Spectrum concentration and market competition. Implications for the use of caps in Mexico. In: Proceedings of the 2015 Conference of Telecommunication, Media and Internet Techno-Economics (CTTE). https://doi.org/10.1109/ctte.2015.7347228
Fuentes Fernández, S. (2023). Time Series: Arima Model Box-Jenkins Methodology (Spanish). Available online: https://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-arima.pdf (accessed on 12 August 2024).
Grand View Research. (2021). Telecom Services Market Size, Share & Trends Analysis Report by Service Type (Mobile Data Services, Machine-To-Machine Services), By Transmission (Wireline, Wireless). Available online: https://www.grandviewresearch.com/industry-analysis/global-telecom-services-market (accessed on 12 August 2024).
Kumari, S., & Muthulakshmi, P. (2024). SARIMA Model: An Efficient Machine Learning Technique for Weather Forecasting. Procedia Computer Science, 235, 656–670. https://doi.org/10.1016/j.procs.2024.04.064
Li, L., Huang, S., Ouyang, Z., Li, N. (2022). A Deep Learning Framework for Non-stationary Time Series Prediction. In: Proceedings of 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA); 20-22 May 2022; Changchun, China. https://doi.org/10.1109/cvidliccea56201.2022.9824863
Lis-Gutiérrez, J.-P. (2013). Market Concentration and Market Stability Measures: An Application for Excel (Market Concentration and Market Stability Measures: An Application for Excel) (Spanish). SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2279769
MathWorks, Inc. (2024). Deep learning time series forecasting (Spanish). Available online: https://la.mathworks.com/help/deeplearning/ug/time-series-forecasting-using-deep-learning.html (accessed on 12 August 2024).
Melnik, A., Shy, O., & Stenbacka, R. (2008). Assessing market dominance. Journal of Economic Behavior & Organization, 68(1), 63–72. https://doi.org/10.1016/j.jebo.2008.03.010
Mills, T. C. (2019). Time Series and Their Features. Applied Time Series Analysis, 1–12. https://doi.org/10.1016/b978-0-12-813117-6.00001-6
MinTIC. (2023). ICT Quarterly Newsletter February 2023 (Spanish). Available online: https://colombiatic.mintic.gov.co/679/articles-274258_archivo_pdf.pdf (accessed on 12 August 2024).
Nekmahmud, Md., & Rahman, S. (2018). Measuring the Competitiveness Factors in Telecommunication Markets. In: Khajeheian, D., Friedrichsen, M., & Mödinger, W. (editors). Competitiveness in Emerging Markets. Springer International Publishing. pp. 339–372.
Ray, S., Lama, A., Mishra, P., et al. (2023). An ARIMA-LSTM model for predicting volatile agricultural price series with random forest technique. Applied Soft Computing, 149, 110939. https://doi.org/10.1016/j.asoc.2023.110939

Wang, C. H., & Liu, C. C. (2022). Market competition, technology substitution, and collaborative forecasting for smartphone panels and supplier revenues. Computers & Industrial Engineering, 169, 108295. https://doi.org/10.1016/j.cie.2022.108295

Wang, H., Song, S., Zhang, G., et al. (2023). Predicting daily streamflow with a novel multi-regime switching ARIMA-MS-GARCH model. Journal of Hydrology: Regional Studies, 47, 101374. https://doi.org/10.1016/j.ejrh.2023.101374

Wang, X., Kang, Y., Hyndman, R. J., et al. (2023). Distributed ARIMA models for ultra-long time series. International Journal of Forecasting, 39(3), 1163–1184. https://doi.org/10.1016/j.ijforecast.2022.05.001

Yasmin, S., & Moniruzzaman, Md. (2024). Forecasting of area, production, and yield of jute in Bangladesh using Box-Jenkins ARIMA model. Journal of Agriculture and Food Research, 16, 101203. https://doi.org/10.1016/j.jafr.2024.101203

Zhang, J., Liu, H., Bai, W., et al. (2024). A hybrid approach of wavelet transform, ARIMA and LSTM model for the share price index futures forecasting. The North American Journal of Economics and Finance, 69, 102022. https://doi.org/10.1016/j.najef.2023.102022