Data and Metadata. 2025; 4:769
doi: 10.56294/dm2025769

ORIGINAL

Regression Models for the Analysis of Telecommunications Data in Ecuador

Modelos de Regresión para el Análisis de Datos de Telecomunicaciones en Ecuador

Jaime G. Alvarado[1], Edgar V. Lema[1], Luis A. Cuaical[2], Alexis D. Alvarado[3]

[1]Universidad Técnica del Norte, Facultad en Ciencias Aplicadas. Ibarra, Ecuador.
[2]Universidad Católica del Ecuador, Facultad de Ciencias Exactas y Naturales. Quito, Ecuador.
[3]Universidad Internacional SEK Ecuador, Posgrado. Quito, Ecuador.

Cite as: Alvarado JG, Lema EV, Cuaical LA, Alvarado AD. Regression Models for the Analysis of Telecommunications Data in Ecuador. Data and Metadata. 2025; 4:769. https://doi.org/10.56294/dm2025769

Submitted: 11-05-2024 **Revised:** 14-09-2024 **Accepted:** 29-03-2025 **Published:** 30-03-2025

Editor: Dr. Adrián Alejandro Vitón Castillo

Corresponding Author: Jaime G. Alvarado

ABSTRACT

The research highlights the importance of mathematical models for better planning in both state and private companies, specifically through curvilinear regressions, to forecast future activities of the State Telecommunications Regulation and Control Agency of Ecuador (ARCOTEL), by analyzing variables such as the number of internet service users. The study was based on data preprocessing, which included homogeneity analysis and scale changes. Statistical tests such as the Mann-Kendall Test and the Helmert Test were applied to evaluate trends in time series. Subsequently, the data were fitted from a linear model to a polynomial one. Evaluation metrics included absolute, mean, and quadratic percentage errors, as well as coefficients of determination and correlation. The analysis showed that the sixth-degree polynomial fitting provided an adequate adjustment for the time series, with high correlation coefficients and relatively low absolute and mean percentage errors, suggesting acceptable accuracy between the fitted and actual values. Scaling the data facilitated comparison and analysis, eliminating biases. The research emphasized the importance of effective planning using mathematical models to predict economic activity in companies. The sixth-degree polynomial fitting proved to be effective in representing time series, with low errors and high accuracy. These methods were useful for planning and forecasting in the telecommunications sector, as exemplified by the analysis of ARCOTEL users.

Keywords: Mathematical Models; Curvilinear Regressions; Polynomial Fitting; Data Normalization.

RESUMEN

Se pone como antecedente de la investigación la importancia de los modelos matemáticos para una mejor planificación en empresas estatales y privadas, específicamente mediante regresiones curvilíneas, para pronosticar actividades futuras de la Agencia Estatal de Regulación y Control de las Telecomunicaciones del Ecuador (ARCOTEL) para analizar variables como el número de usuarios de servicios de internet. El estudio se basó en el preprocesamiento de datos, que incluyó análisis de homogeneidad y cambios de escala. Se aplicaron pruebas estadísticas como el Test de Mann-Kendall y la prueba de Helmert para evaluar tendencias en series de tiempo. Luego, se ajustaron los datos desde un modelo lineal hasta un polinomial. Las métricas de evaluación incluyeron errores porcentuales absolutos, medios y cuadráticos, así como coeficientes de determinación y correlación. El análisis mostró que el ajuste polinomial de grado 6 proporcionó un ajuste adecuado de las series de tiempo, con altos coeficientes de correlación, errores porcentuales absolutos y medios relativamente bajos, lo que sugirió una precisión aceptable entre los valores ajustados y los valores reales. El cambio de escala de los datos facilitó la comparación y el análisis, eliminando sesgos. La investigación subrayó la importancia de la planificación efectiva utilizando modelos matemáticos para predecir la actividad económica en empresas. El ajuste polinomial de grado 6 demostró ser efectivo en la representación de series temporales, con bajos errores y alta precisión. Estos métodos resultaron útiles para la planificación y pronóstico en el sector de telecomunicaciones, ejemplificado por el análisis de usuarios de ARCOTEL.

Palabras clave: Modelos Matemáticos; Regresiones Curvilíneas; Ajuste Polinomial; Estandarización de Datos.

INTRODUCTION

A fundamental aspect of effective business planning, whether state-owned or private, lies in forecasting and anticipating economic activity.[1] To achieve this objective, it is essential to collect and store reliable information over time and to use an adjustment model capable of generating estimates close to actual values.[2] Although it is valid to carry out the forecasting study using qualitative methods, we have opted to use quantitative approaches in this specific analysis for several reasons.

The availability of reliable data from the State Agency for the Regulation and Control of Telecommunications (ARCOTEL) covers a monitoring period of 11 years, including measurements of various parameters, such as the number of users of different operators offering internet services using multiple technologies. This data allows a statistical analysis to be carried out, examining the relationship between the dependent variable (the number of users of the different operators) and the independent variable (time).[3] The distribution of data over time suggests that the behavior of the variable follows a pattern as time passes, facilitating the development of a model at the end of the process that allows for the desired forecast, for the time series analysis, trend, cyclical variation, seasonal variation, and random variation are examined.

Although there are methods such as Neural Networks or the application of techniques such as moving averages, exponential smoothing, or the Box-Jenkins method for time series analysis, in this study, we choose to use polynomial regression methods. This choice is based on evaluating the effectiveness of these methods to identify the underlying pattern in time series, motivated by specific reasons.[4]

Polynomial models are simpler and easier to apply and interpret compared to more complex methods, such as neural networks or autoregressive integrated moving average models, which are used by both state and private operators in the telecommunications field. In addition, these models are remarkably flexible and adaptable to a wide variety of data patterns, making them ideal for capturing the inherent complexity in time series.

In time series related to telecommunications, the specific shape of the curve acquires practical relevance and contributes significantly to the interpretation of the observed trend's results. The best model is selected using decision criteria established through an exhaustive evaluation of these methods, thus guaranteeing the choice of the most appropriate approach for the analysis of time series in the context of telecommunications.[5]

METHOD

The methodology used to analyze the time series of active lines in prepaid modalities, based on the data provided by ARCOTEL, was developed systematically. The process was carried out following these stages:

Data pre-processing

This stage was crucial to guarantee the availability and reliability of the data. A descriptive analysis of the ARCOTEL database was carried out to identify possible anomalous data, ensuring that the data used in the study was reliable and error-free.[6]

Homogeneity analysis with time series

Non-parametric statistical tests were used to identify errors and determine the homogeneity of the series. The Mann-Kendall test and the Helmert statistical test were used.[7]

Mann-Kendall test

This is a non-parametric statistical test used to identify trends in time series. It assesses the existence of a monotonic relationship between the values of the time series and time without assuming a specific distribution of the data. The test calculates the variance (V) and compares it with a threshold to determine the significance of the observed trend. It is beneficial for detecting systematic increases or decreases over time, which makes it ideal for time series studies with hydrological and climatic data where one wishes to identify trends in observed data.[8] Furthermore, the Mann-Kendall test is robust against data with non-normal distributions and can handle time series with missing values, increasing its applicability in various scientific fields.[9]

Helmert statistical test

This test detects structural changes in the time series. It determines if there is a linear relationship between the values of the number of users of the different operators over time. The series is classified as homogeneous or non-homogeneous according to the difference between the sum of the signs and the square root. The Helmert test is used in Gauss-Helmert models for least squares estimation, providing a robust basis for statistical analysis in prediction processes. It is also used in contrast analysis, facilitating the evaluation of theoretical predictions about differences between group means in empirical data.[10]

Data Scaling

Instead of traditional standardization, the data was scaled to 0 and 1. This process, known as Min-Max normalization, is crucial when applying least squares methods to polynomials of degrees more significant than 6. The main reason for this is that, when dealing with tremendous values, the coefficients in polynomials raised to high powers can become extremely large, which makes calculation difficult and can lead to problems with numerical precision. Transforming both the independent variable (time) and the dependent variable (users) to a scale between 0 and 1 ensures that all values remain within a manageable range, facilitating comparison and analysis.[11]

Polynomial Data Fitting

Polynomial data fitting captures the non-linear relationship between the independent variable (time) and the dependent variable (users). This process involves finding the coefficients of a polynomial equation that best fits the observed data. The objective is to minimize the sum of the squares of the differences between the observed values and those predicted by the polynomial model.

The polynomial fit can involve various degrees of the polynomial, from a simple linear fit to polynomials of a higher degree. This study considered models ranging from a linear fit to a 6th-degree polynomial to obtain an adjusted representation of the time series. As the degree of the polynomial increases, the model's ability to adapt to the complexities of the data increases; however, with the risk of overfitting, where the model adjusts too much to the specific variations of the training data, losing its predictive capacity for new data.

The polynomial fitting process uses the least squares method, which looks for the coefficients that minimize the sum of the squares of the differences between the observed and predicted values. This method is fundamental in polynomial regression, as it allows the coefficients of the polynomial equation that best describes the relationship between the variables to be determined.[12]

Fitting Analysis

Several key metrics were used to evaluate the fit quality and the polynomial model's predictive capacity. The Mean Absolute Percentage Error (MAPE) measures the average absolute error as a percentage of the actual values, useful for comparing different models or data sets. The Mean Absolute Error (MAE) calculates the average of the absolute differences between the actual and predicted values, providing a direct measure of accuracy. The Mean Squared Error (MSE) promotes the squares of the differences between the actual and predicted values, penalizing more significant errors more, and its square root, the Root Mean Squared Error (RMSE), allows errors to be interpreted on the same scale as the original values, facilitating comparison.[13] Furthermore, the Coefficient of Determination (R^2) indicates the proportion of the variance in the dependent data that is predictable from the independent variables. In contrast, the Coefficient of Correlation (R) measures the strength and direction of the linear relationship between two variables, providing a complete evaluation of the effectiveness of polynomial modeling in time series analysis.[14]

RESULTS AND DISCUSSION

Data pre-processing

Table 1. Statistical analysis of prepaid and postpaid operators

Descriptive statistics	Postpayment			Prepaid		
	CONECEL	OTECEL	CNT	CONECEL	OTECEL	CNT
Average	7 443 458	3 521 820	114 1387	2 091 800	1 028 243	321 702
Standard error	123 193	34 457	76 564	40 402	21 616	15 041
Median	6 757 632	3 405 847	546 729	2 216 863	1 159 541	325 673
Standard deviation	1 606 241	449 270	998 272	526 776	281 834	196 113
Variance	2,5800E+12	2,0184E+11	9,9655E+11	2,7749E+11	7,9430E+10	3,8460E+10
Kurtosis	-2	-1	-2	0	-1	-1

https://doi.org/10.56294/dm2025769

Coefficient of asymmetry	0	0	0	-1	-1	0
Range	4 598 872	1 602 667	2 598 440	1 751 543	898 193	642 217
Minimum	53 27 336	2 611 348	121 410	928 531	468 235	48 961
Maximum	9 92 62 08	4 214 015	2 719 850	2 680 074	1 366 428	691 178
Sum	1 265 387 891	5 987 09 400	19 4035 844	355 605 990	174 801 323	54 689 338
Count	170	170	170	170	170	170

Data pre-processing was applied to active line users in both modalities (prepaid and postpaid) and to the three operators, CONECEL, OTECEL, and CNT. The analysis was based on the descriptive statistics summarized in the following table.

Descriptive Statistical Analysis

For the postpaid modality, CONECEL has the highest average number of users, followed by OTECEL and then CNT, which has significantly fewer users in comparison. The variability in the number of users is more significant in CONECEL, indicating a wider dispersion of the data around the mean. Regarding kurtosis and asymmetry, the distributions for the three operators are flatter than expected in a normal distribution and are symmetrical. The range of users is also more excellent in CONECEL, showing the most significant variability between the minimum and maximum number of users. For the prepaid modality, CONECEL again leads with the highest average number of users, followed by OTECEL and CNT. Variability is also more significant in CONECEL, suggesting a more excellent dispersion in the data. The distributions in CONECEL and OTECEL are flatter and tend towards the left, while CNTs are more symmetrical. The range of prepaid users is more excellent in CONECEL, which indicates a higher variability.

Interpretation for Polynomial Fitting

The high variability observed in CONECEL suggests that any polynomial fitting model must be able to capture this dispersion to be effective. The lower variability in OTECEL and CNT implies that their models could be more straightforward. Differences in central tendencies, such as the mean and median, indicate possible biases in the data that must be considered during modeling. The flat and asymmetric distributions in some operators suggest the presence of extreme values, which can influence the choice of the degree of the polynomial to avoid overfitting or underfitting. This analysis provides a solid basis for applying polynomial models, ensuring that the key characteristics of the data are adequately considered to achieve an accurate and practical fit.

Homogeneity Analysis with Time Series

Mann-Kendall test

The homogeneity analysis used the Mann-Kendall test to evaluate significant trends in the time series of the operators CONECEL, OTECEL, and CNT in the prepaid and postpaid modalities. As indicated above, the Mann-Kendall test is a non-parametric test used to detect trends in a time series and is based on the following formula:

$$V = \frac{S-1}{\sqrt{\frac{n(n-1)(2n+5)}{18}}}$$

Where:

$$S = T - I$$

T is the number of significant events of the value being analyzed in the time series.

I is the number of minor events of the value being analyzed in the time series.

Below is the table of results from the Mann-Kendall test.

From the table analysis, the level of significance used was 10 % (alpha = 0,1) for each variable. The assigned critical value (V_{crit}=1,28) was used to determine if the absolute value of the test statistic (V= -0,0376) is less than the critical value ($|V| < |V_{crit}|$). The results indicate that the Mann-Kendall test applied to the time series of users of the operators CONECEL, OTECEL, and CNT in the prepaid and postpaid modalities, as well as the total number of mobile phone users, has shown that the series are homogeneous. This indicates that no significant trends were found in the data, which is relevant for planning and predicting economic activity in the telecommunications sector.

Table 2. Mann Kendall test on prepaid and postpaid data

Mann-kendall test	Prepaid CONECEL	Prepaid OTECEL	Prepaid CNT	Postpay CONECEL	Postpay OTECEL	Postpay CNT
Alfa	10 %	10 %	10 %	10 %	10 %	10 %
V_{crit}	1,28	1,28	1,28	1,28	1,28	1,28
I	120	120	120	120	119	120
T	120	120	120	120	119	120
S	0	0	0	0	0	0
V	-0,0376	-0,0376	-0,0376	-0,0376	-0,0376	-0,0376
$\|V\|<\|V_{crit}\|$	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous

Helmert statistical test

To corroborate the previous test, we proceeded with the Helmert statistical test, which was much easier than the Mann Kendal Test and consisted of analyzing the sign of the deviations of each event in the series concerning its mean value, in which if a deviation of a sure sign is followed by another of the same sign a sequence S1 is created, in contrast, if a deviation is followed by another of the opposite sign it is registered as a change C, and is compared with the square root of (n-1), to classify the series as homogeneous or non-homogeneous, according to the formulas indicated below:

$$S1 - C > \sqrt{n-1} \quad => SERIE\ HOMOGENEA$$
$$S1 - C < \sqrt{n-1} \quad => SERIE\ NO\ HOMOGENEA$$

The results of this process are shown in the following table.

Table 3. Helmert's statistical test of the data in the prepaid and postpaid modalities

Average parameters	Prepaid CONECEL	Prepaid OTECEL	Prepaid CNT	Postpay CONECEL	Postpay OTECEL	Postpay CNT
S1=	13	12	14	14	14	13
C=	2	3	1	1	1	2
	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous

From the table analysis, it can be seen that the Helmert test compares the annual values of the data in the prepaid and postpaid modalities with an average value calculated to determine if significant changes indicate breaks in the structure of the series. The results showed that most of the years analyzed are homogeneous; they do not present substantial structural changes. This is evident in the tabulated values, where it can be seen that the majority of the years show an 'S,' indicating stability in the series. Therefore, the series is considered homogeneous because the values of the test remain consistent, as can be seen in the majority of the years. The few cases marked with 'C' (change) indicate years with significant variations but not enough to alter the overall homogeneity of the series. The changes are few and within the limits of a probable error.

Relationship with the Mann-Kendall test

The Helmert and Mann-Kendall tests work together to provide a complete view of the time series structure. While the Mann-Kendall test identifies general (monotonic) trends in the time series, the Helmert test focuses on detecting specific structural changes that could indicate breaks in the series.

In this study, both tests indicated homogeneity in the time series analyzed. This means that the series not only has no significant trends (according to the Mann-Kendall test) but also maintains a consistent structure over time (according to the Helmert test). The confirmation of homogeneity by both tests ensures that the time series are stable and suitable for further analysis.

Contribution to Polynomial Fitting

The homogeneity detected by the Helmert test is crucial for polynomial fitting. The structural stability of the time series allows polynomial models to be applied with confidence, as no sudden changes or hidden trends are expected that could distort the results. Polynomial models can thus capture the underlying relationship between time and the number of users of operators' services, providing accurate and reliable estimates.

In summary, the Helmert test confirms that the time series are suitable for polynomial adjustment, validating their use in planning and predicting trends in the telecommunications sector. The time series' stability and homogeneity guarantee that the adjusted polynomial models will faithfully reflect the dynamics of the data, contributing to better strategic decision-making.

Data scaling

The Min-Max data normalization process in the study is a common and valuable practice for comparing variables on different scales. By transforming the variables to a scale between 0 and 1, a relative comparison between the various payment methods and operators is achieved without the absolute differences in their magnitudes.

This normalization of data allows us to eliminate biases caused by differences in the units of measurement and ranges of the variables, which facilitates the comparison and analysis of the data. By assigning the value 0 to the minimum and the value 1 to the maximum in each payment method, a typical frame of reference is established to evaluate the behavior of users about time and the different payment methods, as we can see in the following table.

For the present analysis, the time variable was assigned 0 to the first value and 1 to the 5072 days corresponding to the last value; this is for the 2 modalities, prepaid and postpaid.

For the user variable, the maximum values for each modality were assigned 1, that is 99,262,080 for prepaid and 26,800,740 for postpaid. Based on these, the table for each modality was completed, as indicated below.

Table 4. Standardization of data in the prepaid and postpaid modalities

Time	Postpayment			Prepaid		
	CONECEL	OTECEL	CNT	CONECEL	OTECEL	CNT
0	0,34646	0,17611	0,01924	0,72490	0,26702	0,02536
0,01	0,34963	0,17620	0,02439	0,73725	0,26308	0,02418
0,01	0,35282	0,17471	0,02461	0,74657	0,26406	0,02418
0,02	0,35667	0,20377	0,02461	0,75324	0,26501	0,02418
0,02	0,35912	0,20559	0,02390	0,76038	0,26503	0,02363
0,03	0,36447	0,17723	0,02240	0,76806	0,27697	0,02434
0,04	0,36782	0,17967	0,02245	0,77334	0,27927	0,02541
0,04	0,37165	0,18237	0,02257	0,77879	0,28401	0,02589
0,97	0,80482	0,45424	0,11493	0,67487	0,41559	0,26618
0,98	0,80962	0,45567	0,11399	0,67805	0,41930	0,26774
0,98	0,81349	0,45691	0,11384	0,68086	0,42219	0,26919
0,99	0,81775	0,45866	0,11389	0,68184	0,42347	0,27075
0,99	0,82250	0,46050	0,11392	0,68741	0,42453	0,27229
1	0,82699	0,46312	0,11365	0,69080	0,41987	0,27401

This treatment ensures that all values remain within a manageable range, facilitating comparison and analysis. A standard frame of reference is established by transforming the independent variable (time) and the dependent variable (users) to a scale between 0 and 1. This eliminates biases caused by differences in the units of measurement and ranges of the variables, allowing for an accurate and consistent evaluation of user behavior over time and between different payment methods.

Data normalization facilitates the application of polynomial models since keeping the values within a uniform range reduces the risk of obtaining substantial coefficients that can lead to numerical precision problems. This is especially important when working with polynomials of degree greater than 6. Standardization ensures that the model fits the data more efficiently and accurately, faithfully reflecting trends and patterns in the time series.

Polynomial Data Fitting

A polynomial approximation with least squares was used to fit the data in the study. The method consists of asking that the sum of the calculated distances between the value of the approximation function p(xi) and the computed value of the function f(xi) be minimal, that is:

$$\sum_{i=1}^{m}(p(x_i) - f(x_i)) = \sum_{i=1}^{m} d_i = \text{mínimo}$$

If we assume a linear regression with, we square the distance to avoid derivability problems:

$$\sum^{m}(a_o + a_1 x_i - f(x_i))^2 = \sum^{m} {d_i}^2 = \text{mínimo}$$

To determine the polynomial, we partially derive with respect to the two variables: First we derive with respect to a_0:

$$\frac{\partial}{\partial a_o}\sum_{i=1}^{m}(a_o + a_1 x_i - f(x_i))^2 = 0$$

Simplification:

$$\sum_{i=1}^{m}(a_o + a_1 x_i - f(x_i)) = 0$$

Applying the properties of summations:

$$na_0 + a_1 \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} f(x_i)$$

Then we derive with respect to a_1

$$\frac{\partial}{\partial a_1}\sum_{i=1}^{m}(a_o + a_1 x_i - f(x_i))^2 = 0$$

Simplification:

$$\sum_{i=1}^{m}(a_o x_i + a_1 {x_i}^2 - f(x_i)x_i) = 0$$

Applying the properties of summations:

$$a_0 \sum_{i=1}^{m} f(x_i) + a_1 \sum_{i=1}^{m} {x_i}^2 = \sum_{i=1}^{m} f(x_i)\, x_i$$

The resulting system is:

$$ma_0 + a_1 \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} f(x_i)$$
$$a_0 \sum_{i=1}^{m} f(x_i) + a_1 \sum_{i=1}^{m} {x_i}^2 = \sum_{i=1}^{m} f(x_i)\, x_i$$

When solving, the first degree polynomial is found: $p(x)=a_o+ a_1x$

In the present investigation, a good fit was not obtained with a first-degree polynomial; the same polynomial regression method was used until a polynomial of degree, n, for example, was reached. The procedure is based on minimizing the function:

https://doi.org/10.56294/dm2025769

$$\sum_{i=1}^{m}[a_o + a_1 x_i + a_2 x_i^2 + \cdots + a_n x_i^n - f(x_i)]^2 = 0$$

If we partially differentiate with respect to the coefficients with , and set each of the derivatives equal to zero, we obtain the following system of equations.

$$\begin{aligned}
ma_0 \quad + \quad a_1\sum x \quad + \quad a_2\sum x^2 \ + \cdots + \ a_n\sum x^n &= \sum y \\
a_0\sum x \ + \ a_1\sum x^2 \ + \ a_2\sum x^3 \ + \cdots + \ a_n\sum x^{n+1} &= \sum xy \\
a_0\sum x^2 \ + \ a_1\sum x^3 \ + \ a_2\sum x^4 \ + \cdots + \ a_n\sum x^{n+2} &= \sum x^2 y \\
a_0\sum x^n + a_1\sum x^{n+1} + a_2\sum x^{n+2} + \cdots + a_n\sum x^{n+n} &= \sum x^n y
\end{aligned}$$

Where the subscripts of the variables and the summations have been omitted to simplify the writing.

In the present investigation, the non-linear analysis was carried out on the time series of active lines, using polynomials from n=1 to n=6.

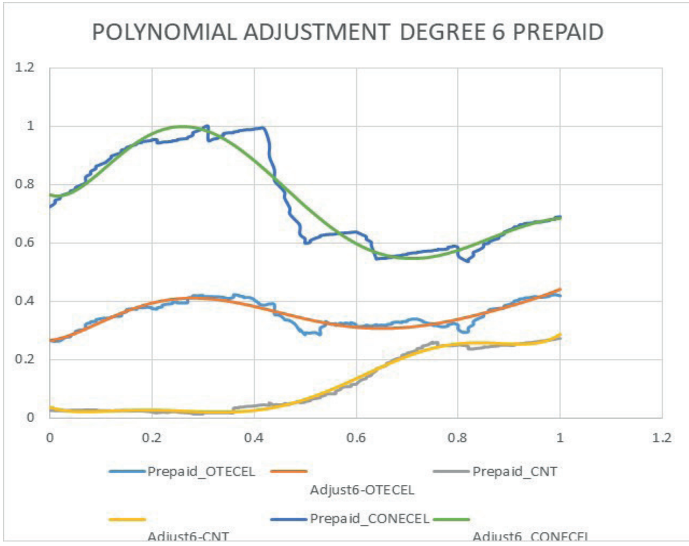

Figure 1. Polynomial adjustment of data in the prepaid modality

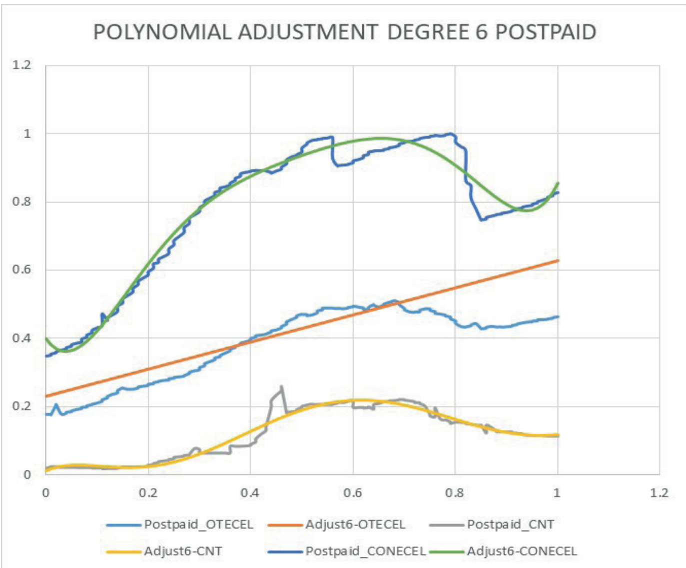

Figure 2. Polynomial fit of the post-paid data

The degree of the polynomial was increased, and the best graphical fit was obtained with a degree 6 polynomial. This analysis was based on the adequacy of the fit to the data, the interpretation of the results, and the subsequent analysis of the error, which is indicated below.

The degree 6 polynomial fit showed a high correlation with the observed data, indicating a good fit of the model. The absolute and mean percentage errors were relatively low, suggesting an acceptable accuracy between the adjusted and actual values. This accuracy and the good fit were achieved by transforming the data to a manageable scale between 0 and 1, which facilitated analysis and comparison without introducing biases due to the original magnitudes of the data.

Fit analysis

The table presents the analysis of the polynomial fit of the data. The coefficients obtained from the fit of the polynomial model to the different time series of the prepaid and postpaid modalities of the different operators are summarized in the following table.

Table 5. Coefficients, correlation and average errors in the prepaid and postpaid modalities

n	Prepaid -CONECEL	Prepaid -OTECEL	Prepaid -CNT	Post-payment -CONECEL	Post-payment -OTECEL	Post-payment -CNT
a0	0,7652	0,2667	0,036	0,036	0,1663	0,011
a1	-0,5894	0,2236	-0,5954	-0,5954	0,8163	0,6641
a2	22,286	6,984	7,4723	7,4723	-5,0621	-8,5855
a3	-98,698	-35,314	-37,699	-37,699	23,535	40,534
a4	163,27	62,838	85,142	85,142	-41,972	-74,146
a5	-117,31	-48,45	-83,73	-83,73	30,362	58,229
a6	30,961	13,893	29,659	29,659	-7,393	-16,589
R^2	0,9349	1	0,9917	0,9917	0,9946	0,944
MAPE	3,84	2,3	13,69	3,42	2,3	12,6
MAE	0,0286	0,008	0,007	0,026	0,008	0,011
MSE	0,002	1,00E-04	8,30E-05	0,001	1,00E-04	3,00E-04

Data analysis shows that the 6th-degree polynomial fit provides a robust model for the time series of prepaid and postpaid users for the different operators. The high accuracy of the model is reflected in the low MAPE, MAE, and MSE values, as well as in the high values, which suggests that the model is adequate for describing and predicting the trends observed in the data.

The coefficients of determination (indicated in the table) are high, which suggests a good fit of the polynomial model to the data. These values range from 0,9349 to 1, indicating that the 6th-degree polynomial model can explain between 93,49 % and 100 % of the variability in the observed data.

The lower values of the Mean Absolute Percentage Error (MAPE) indicate a higher precision of the model. In this study, the values range from 2,3 % to 13,69 %, which suggests that the model has an acceptable percentage precision in most time series.

The low ((Mean Absolute Error) MAE indicates a minor difference between the adjusted and actual values. The values obtained vary between 0,007 and 0,0286, suggesting a good fit of the model to the data for a 6th-degree polynomial.

In this indicator, on the other hand, lower values are observed for the MAE (Mean Absolute Error), which indicates a smaller difference between the adjusted values and the real values. This suggests a good fit of the model to the data, which is also the best fit for a degree 6 polynomial.

The values of the (Mean Squared Error) MSE, which range between 1,00E-04 and 0,002, are relatively low. This indicates a smaller discrepancy between the adjusted values and the real values, which again confirms a good fit for a degree 6 polynomial.

Return prediction of future data

For the prediction of future values, having the adjustment formulas in terms of the coefficients indicated in table 5, we have the formulas that allow us to predict future values. As an example for the operator CONECEL in the prepaid payment modality, the adjustment formula is: $y=0,7652-0,5894\ x+22,286\ x^2-98,698\ x^3+163,27\ x^4-117,31\ x^5+30,961\ x^6$.

Predictions obtained using the formulas allow the operator to anticipate future demand for prepaid or postpaid services, thus facilitating strategic and operational planning. This includes resource management, infrastructure planning, and the development of marketing strategies aimed at satisfying projected demand.

https://doi.org/10.56294/dm2025769

CONCLUSIONS

This article highlights the importance of effective planning in forecasting the number of users, which directly impacts the economic factors for both state-owned and private companies. Time series analysis in telecommunications requires robust techniques to detect trends and structural changes in the data. Statistical methodologies such as the Mann-Kendall test and the Helmert contrast provide a solid basis for interpreting and predicting temporal patterns, facilitating the forecasting of the evolution of the number of internet service users.

Data standardization improves comparison and analysis, allowing for accurate evaluation of company planning and forecasting, particularly in the case of ARCOTEL and its analysis of telecommunications users. Various mathematical models, including curvilinear regressions, have proven fundamental in this process. The study results indicate that the 6th-degree polynomial fit adequately represents the time series, evidenced by high correlation coefficients that suggest a good fit of the model. The absolute percentage errors and mean squares are relatively low, indicating an acceptable precision between the adjusted and actual values.

BIBLIOGRAPHIC REFERENCES

1. Bala R, Ojha DB. Modeling and forecasting Internet traffic using hybrid ARIMA-SVM approach. J Netw Comput Appl. 2021;182:103026. https://doi.org/10.1016/j.jnca.2021.103026

2. Li Y, Xie Z. Network traffic prediction based on a new model with cross-bidirectional long short-term memory. IEEE Trans Netw Serv Manag. 2019;16(3):801-811. https://doi.org/10.1109/TNSM.2019.2902795

3. Smith, J. A., & Lee, C. H. (2019). Analyzing the impact of socio-economic variables on public health outcomes through regression analysis. Public Health Journal, 45(3), 310-325. https://doi.org/10.1093/pubmed/fdy091

4. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 2003;50(3):159-175. https://doi.org/10.1016/S0925-2312(01)00702-0

5. Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control. John Wiley & Sons; 2015. https://doi.org/10.1002/9781118619193

6. Hodge VJ, Austin J. A survey of outlier detection methodologies. Artificial Intelligence Review. 2004;22(2):85-126. https://doi.org/10.1023/B:AIRE.0000045502.10941.a9

7. Helsel DR, Frans LM. The regional Kendall test for trend. Environmental Science & Technology. 2006;40(13):4066-4073. https://doi.org/10.1021/es051650b

8. O'Connell RH, Coberly WT. The Helmert contrast as a trend test. Biometrics. 1975;31(2):301-309. https://doi.org/10.2307/2529441

9. Hirsch RM, Slack JR, Smith RA. Techniques of trend analysis for monthly water quality data. Water Resources Research. 1982;18(1):107-121. https://doi.org/10.1029/WR018i001p00107

10. Meng Q, Lee JH, Zhu S. Application of the Gauss-Helmert model for trend analysis in hydrological time series. Journal of Hydrology. 2018;558:85-98. https://doi.org/10.1016/j.jhydrol.2018.01.015

11. Han J, Pei J, Kamber M. Data Mining: Concepts and Techniques. Morgan Kaufmann; 2011. https://doi.org/10.1016/C2009-0-61819-5

12. Draper NR, Smith H. Applied Regression Analysis. John Wiley & Sons; 1998. https://doi.org/10.1002/9781118625590

13. Makridakis S, Wheelwright SC, Hyndman RJ. Forecasting Methods and Applications. John Wiley & Sons; 1998. https://doi.org/10.1002/9781119263678

14. Nagelkerke NJD. A note on a general definition of the coefficient of determination. Biometrika. 1991;78(3):691-692. https://doi.org/10.1093/biomet/78.3.691

FINANCING

None.

CONFLICT OF INTEREST

None.

AUTHORSHIP CONTRIBUTION

Conceptualization: Luis Cuaical.
Data curation: Edgar V. Lema.
Formal analysis: Jaime Alvarado.
Research: Jaime Alvarado.
Methodology: Jaime Alvarado.
Project administration: Jaime Alvarado.
Software: Alexis D. Alvarado.
Supervision: Jaime Alvarado.
Validation: Jaime Alvarado.
Visualization: Luis Cuaical.
Writing - original draft: Jaime Alvarado.
Writing - review and editing: Jaime Alvarado.