# The frontier of jurisdictional excellence: new perspectives of economics given digital transformation in Brazil's ecosystem

Aline Araujo Perini[1]*; Ricardo Francisco Esposto[2]; Carlos A. Grespan Bonacim[3]

[1] Prof. Dr. Finance and Controllership, Digital Business, International Affairs. St Cezira Giovanoni Moretti, 580 – Santa Rosa; 13.414-157 Piracicaba, SP, Brazil; orcid.org/0000-0002-3058-0224
[2] Pecege, Advisor Professor. St Alexandre Herculano, 120 – Vila Monteiro; 13418-445 Piracicaba, SP, Brazil;
[3] Professor at the School of Economics, Administration and Accounting at Ribeirao Preto (FEARP) - University of Sao Paulo. Av. Bandeirantes, 3900 - Monte Alegre; 14040-905 Ribeirão Preto, SP, Brazil; orcid.org/0000-0003-0347-9419

*corresponding author: alineperini@alumni.usp.br

## Abstract

The objective was to understand phenomena related to the efficiency function of the Brazilian justice system by estimating 16 Generalized Linear Models [GLM]. The Crisp-DM and Rastro-DM methodologies provided a summarized overview of the learning and optimization trajectory. The Global Justice (60) entities include the State Courts of Justice [TJEstaduals] (27), Regional Labor Courts (27) [TRTs], and Federal Regional Courts [TRFs] (6). The initial analysis of four straightforward linear models linked efficiency to time from 2014 to 2021. The significance of the parameters from the TJEstaduals sample led to their selection for the multivariate modeling stage, which included nine variables covering the entire timeframe from 2009 to 2022. In the study of five TJEstaduals models (27), the explanatory variables new cases (ch) and computers per user (inf2) proved significant in all cases. The complete model demonstrated a better alignment with the efficiency phenomenon, free from the effects of multicollinearity or heteroscedasticity. In the detailed component analysis of ch unit features, the most accurate estimate was the ARIMA prediction (0,0,0) with a zero mean, no constant trend, and one lag minor MAPE error, showing no ARCH effects or residual correlation. Using data from the Global model (60), covering the period from 2015 to 2019 without distinguishing between court or justice categories, seven new models were examined, applying efficiency adjustments using dummies for count data. The optimization was achieved with the Zero-Inflated Negative Binomial [ZINB] model, which characterizes the efficiency phenomenon due to the superposition of data inflated with zeros.

**Keywords:** institutional; governance; prevention; long-life learning

## 1. Introduction

The OECD (2024) Global Trends in Government Innovation report promotes human-centric public administrative services principles. The trends identified are: i) future-oriented and co-created public services; ii) digital and innovative foundations for efficient public services; iii) personalized and proactive services for accessibility and inclusion; iv) data-driven public services for better decision-making; and v) public services that create opportunities for public participation. It outlines how public services should reflect value creation amid technological changes, shifting preferences, and an increasingly complex government ecosystem. Innovation is a key tool for strengthening institutions, fostering collaboration, and establishing structural governance, resources, and skills for human-centric public services.

The United Kingdom National Audit Office [NAO] (2025) strives to map best practices and lessons learned across the government to inform and enhance disclosure and compliance policies. To address the backlog in the Crown Court caused by the COVID peak, they examined: i) the scale, nature, and impact of the Crown Court backlog; ii) the interfaces between the Ministry of Justice [MoJ] and His Majesty's Courts & Tribunals Service [HMCTS] to understand the effects of actions taken to reduce the Crown Court backlog; and iii) how the MoJ and HMCTS are attempting to manage the Crown Court backlog. The NAO concludes that: i) the backlog peak likely exacerbated the negative effects of prolonged wait times and also increased the costs of administering justice; ii) the MoJ must collaborate closely with other components of the system to gather intelligence for responsive action; and iii) the Criminal Justice Board [CJB] should explore ways to establish resilience and agility within the system to ensure cases are not unduly delayed.

Alon-Barkat and Shamir (2023), in a counterpoint, amend how the proposed judicial reform in Israel may compromise integrated approaches such as "Health in All Policies" [HiAP], a strategy that requires strong intersectoral articulation to promote well-being in human decision-making. In this sense, governance and the judicial system are essential pillars for political decisions to consider health across all spheres. The attempt to concentrate power in a single decision-making axis threatens the system's complexity, where human institutions

with diverse values, dialogue, and engagement are necessary to ensure evidence-based policies, social justice, and equity.

Alyas et al. (2025), in their paper " An Integrated Solution for Judicial Case Management using Blockchain Technology and Smart Contracts, " demonstrate that a multi-blockchain architecture, combined with smart contracts and decentralized protocol storage, provides a more transparent, scalable, and efficient judicial system, marking a significant advancement in modernization processes. This architecture enables a judicial case management approach aimed at ensuring performance targets, reducing case backlog, supporting complexity analysis for constructing consensus mechanisms, and decreasing system vulnerabilities. It reflects a reality of collaborations with courts and legal institutions, which can yield insights into the system's real-world challenges, practical value, and limitations.

Filgueiras et al. (2025) warned that algorithms, much like institutions, have direct implications for institutions. The authors propose that human centrality should not be abandoned but resignified in the light of new socio-technical arrangements. Instead of resisting automation, one should approach forms of integration that respect ethical values, epistemic diversity, and transparency. This implies developing algorithmic governance models that consider the role of institutions as guarantors of rights, pluralism, and cognitive justice, ensuring that technological mediation remains subordinated to human and social purposes.

Hess (2010) noted that in Brazil, the principle of efficiency in the judicial system was introduced through the reform of the Constitutional judicial system established by [EC45/04]. The reform provided for access, procedural time, digital transformation of procedures, document digitization, and the integration of notary offices and other administrative entities as means to achieve the desired efficiency in public management and jurisdictional provision.

Nissi et al. (2019) studied territorial disparities of efficiency across Italy, under the approach of improving management in the public sector and the importance of the judicial system to maintain peace and coexistence between people and nations. Giancotti et al. (2024) examined factors affecting efficiency in Italy and mapped potential management actions for improvement. In their systematic literature review of 22 publications, they categorized the factors into three main classes: human resources management, judicial processes, and factors internal or external to judicial instances.

Beccaria (1764), in “Dei delitti e delle pene”, did not deal with numerical data but established the foundations of rational justice, proportionality, and prevention. His ideas about the demand feature to structure the judicial system based on logical and equitable principles later influenced the use of data as a tool to evaluate justice. Beccaria was one of the first to suggest that the criminal process should follow rational and predictable criteria, which later inspired systematic data collection and analysis methods to measure and improve the efficiency and equity of judicial systems.

The general objective was to understand the phenomena related to the efficiency of the Justice System in Brazil. Specific objectives include: i) investigating the relationship between digital transformation and judicial efficiency by considering various structural, technological, and human resource variables; ii) understanding and using General Linear Models [GLMs], focusing on counting models (e.g., Poisson and Zero-Inflated Negative Binomial [ZINB]) to capture patterns of overdispersion and excess zeros in the data of the Justice System; iii) exploring the applicability of statistical models in different sectors beyond the judiciary, analyzing how these methodologies can help optimize processes and guide public policies in areas such as Health, Education, and Transportation.

## 2. Material and Methods

Schiavon (2021) stated that studies on the efficiency of the judiciary have become increasingly frequent due to the dynamics of digital transformation in both developed and developing countries, especially in emerging markets like Brazil. Quality metrics reflect the course of transformations, proposals for administrative reform, and radical governance under performance dimensions, work structures, consolidation of positions, functions and bonuses, arrangements of institutional networks, career guidelines, and the civil servant statute.

Schiavon (2021) found evidence in Brazil´s Court of Justice of the State of Minas Gerais [TJMG], the second largest in terms of the number of judges, only behind the Court of the State of São Paulo [TJSP], that digital transformation and technological integration are associated with increased productivity of magistrates, compliance rates with demands, as well as the number of cases resolved and the number of new cases, which are objective criteria related to the efficiency of the

TJMG. Schiavon (2021) suggests the inclusion of variables from the second instance, in addition to further studies on the topic, as a proposal for managing the judiciary's image under the framework of contract theory rather than criminal litigation. According to Schiavon (2021), the digital transformation of the judicial ecosystem is the key to economic and social transformation.

Souza and Guimarães (2018) examined the relationships among resources, innovation, and performance in 24 Brazilian labor courts from 2003 to 2013 using data envelopment analysis and stochastic frontier analysis. The stochastic model indicated that the size of the court and investment in staff training were key factors in explaining the variation in court efficiency. The results revealed that improvements in court performance are more related to the adoption of innovations than to variations in technical efficiency, meaning technological adoption and digital transformation.

Tenório and Souza (2021) analyzed the following factors: conciliation, technology, judicial organization, budget composition, and demand. Using explanatory-deductive analysis, they related these factors to the efficiency of Brazilian state Courts of Justice from 2015 to 2019. Although the authors' analytical approach aligns more closely with litigation control theory than with contract management, they regarded cultural oscillation as an external aspect. In analyzing demand as an external factor, two scenarios were established: one involving constant demand and the other featuring demand with variation effects. The number of personnel in technology was significant for both constant and variable demand. Among the internal factors, the number of judicial courts per inhabitant and the collection of both own revenue and litigation fees exhibited a positive correlation with efficiency, indicating that the volume of demand has a meaningful relationship with efficiency.

Kapopoulos and Rizos (2024) examined the impact of judicial efficiency on the economic growth of the Court of Justice of the European Union using a dataset from 2010 to 2018. They estimated a controlled growth equation linked to indicators of productivity and judicial efficiency. The study revealed that operational inefficiencies obstruct economic growth, undermining the Court's ability to protect private contracts and property rights. The endogeneity of model evaluation carries significant policy implications.

Computational thinking is a strategy for solving problems using cognitive skills in the field of computer science. Almeida et al. (2022) propose the systematic mapping of literature (MSL) for learning computational thinking [CP] from 2010 to 2021, selecting 22 studies published in journals and conference proceedings, primarily from the USA, with a central trend emphasizing the significant value of training teachers in CP.

Djiroun et al. (2018) defined an integrated intelligence system that aligns with certified information as a database [DB], which is updated periodically. In the decision-making process, data is typically stored in the DB in the form of multidimensional cubes. This structure can generate new demands for data requests, leading to the creation of new cubes and resulting in a significant number of managed cubes, along with complexity and heterogeneity of data. Djiroun et al. (2018) proposed the approach in two variants: one based on analysis indicators and another based on the known cube. The validation of the approach was carried out using a tool called "Design-Cubes-Query," which implemented the methodology through a case study.

Castro and Balaniuk (2020) published a methodological proposal in the Journal of the Federal Court of Accounts [TCU] that emphasizes understanding the unit's historical learning traits, the training conducted, the planned actions, results, absorption, and the development of intelligence aimed at organizational advancements and knowledge sharing, rather than focusing solely on the generated model. The Rastro-DM methodology is proposed as a complement.

The Cross-industry Standard Process for Data Mining [CRISP-DM], developed by Chapman et al. (2000), plays a significant role in standardization. In contrast, Rastro-DM, as proposed by Castro and Balaniuk (2020), underscores the importance of actionable steps, selection of techniques, training, and the synthesis of knowledge derived from CRISP-DM. DM-Trace denotes the action trajectory, while CRISP-DM is concerned with object distribution. Together, these methodologies provide a comprehensive perspective on lifelong learning.

Christoph Schröer et al. (2021) reviewed the systematic literature on data mining and the CRISP-DM methodology in specialized publications found in the databases of the Institute of Electrical and Electronics Engineers [IEEE], Science Direct, and the Association for Computing Machinery [ACM], 20 years after the initial publication in the field by Chapman et al. (2000). The authors identified six

main phases or stages of the CRISP-DM method related to standardization in a multidisciplinary area. The six stages were: i) business model learning; ii) data understanding; iii) data preparation; iv) modeling; v) evaluation, and vi) distribution

Christoph Schröer et al. (2021) analyzed that the greatest challenge identified in the studies does not account for the distribution or application phase. The most cited phase involved modeling studies, which included the analysis of different models, technology used, construction of accuracy tests, and visualization of results. Figure 1 summarizes the methodological proposal.

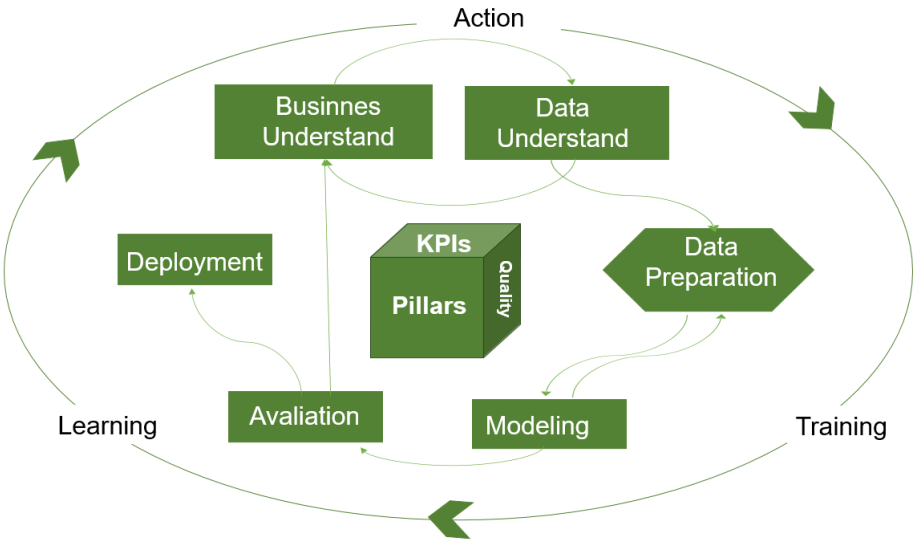


Figure 1. Rastro-DM and Crisp-DM a multidimensional single cube approach
Source : Djiroun et al. (2018), Chapman et al. (2000) e Castro e Balaniuk (2020)

The judicial network is comprised of the national justice system and includes the superior bodies of the Superior Court of Justice [STJ], the Superior Labor Court [TST], the Superior Electoral Court [TSE], and the Superior Military Court [STM]. It consists of six Federal Regional Courts [TRFs], 24 Regional Labor Courts [TRTs], 27 State Courts of Justice [TJEstaduals], 27 Regional Electoral Courts [TREs], four State Courts of Military Justice, and the Military Court of the Union. The latest update from the CNJ [2024] indicates that the database was last updated on August 23, 2023, featuring indicators of the judicial network from January 2011 to December 2022 on an annual basis, comprising 1,314 categorical performance variables. The complete database contains 1,695,060 observations.

In Figure 2, the business model based on the pillars of the digital transformation of the Brazilian judicial system includes society, internal processes, and learning and growth. The three-dimensional cube serves as a framework to meet the objectives of research related to digital transformation, fine-tuning data engineering relevant to justice units, time series by units, and variables aggregated by category. This approach provides multidimensional analytical analysis, detects data behavior, generates algorithms, and enables predictions with accuracy and quality (CNJ, 2021; Heine, 2017).


Figure 2. Pillars of the digital transformation of the judicial network
Source: adapted from CNJ (2021) and Heine (2017)

In analyzing judicial efficiency, variables such as 'expenditure per inhabitant' and 'workforce' were selected due to their relevance demonstrated in previous studies linking financial and human resources to the operational capacity of public systems. For instance, Tenório and Souza (2021) showed that higher per capita expenditures are associated with improvements in the efficiency of public

services, thereby justifying their inclusion in this study. The findings indicate that the combination of financial allocation rates, human resources, and infrastructure, along with effective workload management, is a crucial determinant for enhancing the efficiency of state courts of justice.

Recognizing the importance and effects of the results, this study adopted a hybrid methodology grounded in the principles of Rastro-DM and CRISP-DM. These approaches are widely recognized among data scientists for their effectiveness in lifelong learning. The proposed methodology can be applied across various sectors, including industry, services, and the public sector, providing a deeper understanding of the elements that make up the ecosystem and identifying opportunities for integrated improvement and lifelong learning.

### 2.1 Simple linear models

During the exploratory data phase, the efficiency ($\widehat{Y}$) as a function of time (X) in years was analyzed, with the initial pilot collection covering the period from 2014 to 2021, totaling 7 years. The four models studied unified the set of Global Justice (60), which includes the set of State Courts of Justice [TJEstaduals] (27), the set of Regional Labor Courts [TRTs] (27), and the set of Federal Regional Courts [TRFs] (6). Fávero and Belfiore (2022) indicated that the expected value is known as the conditional mean, calculated for each observation as a function of time. The problem was modeled using equation (1):

$$Efficiency\ (\widehat{Y}) = f\ (year\ x) \tag{1}$$

The explanatory power of the regression model is represented by the adjustment coefficient R², which can vary between 0 and 1 (0% to 100%). The closer it is to 1 or 100%, the more the variable $\widehat{Y}$ is explained as a function of X, with no residuals. To analyze the explanatory capacity (Fávero et al., 2009; Fávero and Belfiore, 2022), R² is calculated in expression (2):

$$R^2 = \frac{\sum_{i=1}^{7}(\hat{Y}_i - \bar{Y})^2}{i=\sum_{i=1}^{7}((\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{7}(\mu)^2} = \frac{SQR}{SQR+SQU} = \frac{SQR}{SQT} \tag{2}$$

Where: deviation of the values of the regression line in relation to the mean; $(\hat{Y}_i - \bar{Y})$:

μ: is the average of the observed values of Y;

SQR: sum of squares of the regression;

SQT: total sum of squares.

According to Fávero and Belfiore (2022), Pearson's correlation coefficient (ρ) ranges from 1 to -1, with stronger relationships occurring at the extremes and weaker relationships approaching zero. The signal allows for the verification of the directionality of the relationship between the variables. A positive sign indicates a directly proportional relationship, whereas a negative sign indicates an inversely proportional relationship. The coefficient (ρ) is calculated using expression (3):

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x}).(Y_i - \hat{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}.\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \quad (3)$$

According to Fávero and Belfiore (2022), it is essential to assess the overall statistical significance of the estimated model using the F test, where the null and alternative hypotheses are:

H0 = b = 0

H1 = β ≠ 0

To Fávero and Belfiore (2022), the F-test is calculated using expression (4):

$$F = \frac{\frac{R^2}{(k-1)}}{\frac{(1-R^2)}{(n-k)}} \quad (4)$$

Where:

K-1: degree of freedom of the regression;

N- K: Degrees of freedom for waste.

According to Fávero and Belfiore (2022), the Shapiro-Francia test can be applied to samples of size 5 ≤ n ≤ 5,000, under the following hypotheses:

H0: the sample comes from a population with a Normal distribution (μ, σ)

H1: the sample does not come from a population with a Normal distribution (μ, σ)

The Shapiro-Francia statistic is provided in expression (5):

$$W'cal = \frac{\left[\sum_{i=1}^{n} mi \,.x_i\right]^2}{\left[\sum_{i=1}^{n} mi^2 . \sum_{i=1}^{n}(x_i - \bar{x})^2\right]}, \quad . \quad (5)$$

where:

$x_{(i)}$ are statistics of the order i of the ordered sample; mi is the approximate value of the ith observation, approximated by $m_i = {}^{\Phi-1}[i/(n+1)]$; ${}^{\Phi-1}$ is the inverse of the normal distribution.

For Fávero and Belfiore (2022), the study of statistical distribution behavior is not new, primarily beginning in the early nineteenth century with normal distribution modeling. Although recent GLM studies have focused on specific cases, these include:

- Simple and multivariate linear models with Box-Cox transformation;
- Logistic and multinomial regression models;
- Poisson and negative binomial regression models with count data.

Salazar-Salazar et al. (2024) defined big data analysis systems as modern software platforms with descriptive, predictive, or prescriptive purposes, made viable by the availability of massive data sources, both internal and external to the organization. The authors point out that there is limited comparative literature regarding development rigor-oriented plan methodologies versus agile, demand-oriented methodologies. To address this gap, they provide a comparative review of CRISP-DM, the primary methodology focused on development rigor, and Team Data Science, a proprietary agile methodology that employs Scrum-XP workflows and practices, offering a theoretical foundation for understanding the differences between development rigor and agile management.

### 2.2 Multivariate generalized linear models

The TJEstaduals database (27) is part of the set of 26 State Courts of Justice plus the Federal District [DF]. It was updated with the entire historical series available from 2009 to 2022, totaling 14 years and including 8 explanatory variables: new cases [ch], Justice expenditure per 100,000 inhabitants [g7]; IT workforce total [f5], second-degree judgment [public2], number of computers per user [inf2], total area ($m^2$) [mtotal], total staff of the judicial schools and magistrates [tpesc], and the workforce of the judicial schools with the total [f6].

The explanatory variables are connected to the three pillars of digital transformation: society, internal processes, and learning and growth, aligning the explanatory variables of both the internal and external environments (Table 1).

Table 1. Variables for the study of the TJEstaduals system

| Eff | Ch | g7 | f5 | apublic2 | inf2 | mtotal | TPESC | f6 |
|---|---|---|---|---|---|---|---|---|
| Efficiency (Y) | New cases per 100,000 inhab. | Total Justice Expenditure per inhab. (R$) | IT Work force with total | Trials in the 2nd instanc | Num. of Comp. per User | Total area (m2) | Total Staff of the Judicial and Magistrate Schools | Judicial Schools Workforce with total |
| Pillars Digital Transformation | extern | extern | intern | intern | intern | intern | extern | extern |
| | Society | | | Internal process | | | Learning and growth | |

Source: original research data

Sttokel-Walker and Noorden (2023) highlight the significance of ChatGPT and Generative Artificial Intelligence [AGI] for advancements in science. They propose that the integrated solutions of these technologies can enhance various aspects, including the quality of the produced texts, research methodology, user experience, and the time taken to analyze technical reports. However, scientists experience a blend of excitement and apprehension regarding these advances. In particular, they may be worried about ethical concerns such as algorithmic bias and the potential effects on researchers' autonomy and roles.

### 2.3 Multicollinearity diagnosis of multivariate models

The first step involves analyzing the high correlations among the explanatory variables using a simple correlation matrix. The second step is examining the determinant of the det(X'X) matrix, where very low values may indicate high correlations between the explanatory variables, calling into question the independence of these variables and their convergence to the same factor or area of intersection. According to Vasconcelos and Alves (2000), each regression expression has an $R^2$, and the analysis of Tolerance along with the Variance Inflation Factor [VIF] can be applied in equations (6) and (7):

$$Tolerance = 1 - R_k^2 \quad (6)$$

$$VIF = \frac{1}{Tolerance} \quad (7)$$

Fávero and Belfiore (2022) state that a high level of multicollinearity is indicated by a VIF above 10. If tolerance is very low, resulting in a high VIF statistic, multicollinearity issues may arise. A low tolerance indicates that the explanatory variable acts similarly to the dependent variable, sharing a significant amount of variance with other explanatory variables.

### 2.4 Causes of heteroscedasticity of multivariate models

Fávero and Belfiore (2022) stated that the non-constancy of the variance of the residuals along the explanatory variable can lead to heteroscedasticity, which is correlated with errors. The variable X is visualized as the formation of a cone that narrows as X increases or decreases. Omitting relevant variables can generate error terms in the model. However, causation can be attributed in learning-and-error models as predictions converge, forming a cone or resulting in an increase in discretionary income or the presence of outliers. The Breusch-Pagan/Cook-Weisberg test detects heteroscedasticity, which is indicated in cases assuming normality of the residuals, in which:

H0: the variance of the error terms is constant (homoscedastic errors)

H1: the variance of error terms is not constant, being a function of one or more explanatory variables (heterocedaceous errors).

### 2.5 Regression model with Box-Cox transformation

The Box-Cox transformation regression model is given in expression (8):

$$\frac{\widehat{Y_i^{\lambda}}-1}{\lambda} = \alpha + \beta_1.X_{1i} + \beta_2.X_{2i} + \cdots \beta_k.X_{ki} \quad (8)$$

where: is the expected value of the dependent variable Y and λ is the Box-Cox transformation parameter.$\hat{Y}$:

### 2.6 Time series

A principal component analysis of the accumulated data series of the variable ch, new cases of the TJEstaduals, was conducted using the Box-Jenkins methodology (ANALISEMACRO, 2024). The Dickey-Fuller stationarity test (1979)

and autocorrelation analysis [ACF], partial autocorrelation [PACF] were applied for estimation, utilizing the autoarima function in RStudio (Hyndman and Khandakar, 2008), analyzing the correlation of the residuals of the estimated models.

- Error statistics analyzed for the predicted time series models were:
- • "Mean Error" [ME]: the average difference between actual and predicted values, as shown in expression (9):

$$Erro_t = X_t - \hat{X}_t \; ; \qquad ME = \frac{\sum_{t=1}^{h} erro_t}{h} \tag{9}$$

- "Mean Absolute Error" [MAE]: average of the realized and predicted absolute difference, in expression (10):

$$MAE = \frac{\sum_{t=1}^{h} erro_t}{h} \tag{10}$$

- Root Mean Square Error [RMSE]: total standard deviation of the sample of the difference between predicted and realized, in equation (11):

$$RMSE = \sqrt{\frac{\sum_{t=1}^{h} (erro_t)^2}{h}} \tag{11}$$

- "Mean Percentage Error" [MPE]: percentage difference of the error, in expression (12):

$$MPE = \frac{\sum_{t=1}^{h} \frac{erro_t}{x_t}}{h} . 100\% \tag{12}$$

- "Mean Absolute Percentage Error" [MAPE]: Absolute percentage difference of the error, as shown in expression (13):

$$MAPE = \frac{\sum_{t=1}^{h} \frac{erro_t}{x_t}}{h} . 100\% \tag{13}$$

- "Theil Inequality Coefficient, Theil's U" [TIC]: reflects the degree of adjustment of the forecast; the lower the value, the better, with zero being ideal, as expressed in equation (14):

$$TIC = \sqrt{\frac{\sum_{t=1}^{h} \left( \frac{\widehat{x_{t+1}} - x_{t+1}}{x_t} \right)^2}{\sum_{t=1}^{h} \left( \frac{x_{t+1} - x_t}{x_t} \right)^2}} \tag{14}$$

- "First-Order Autocorrelation Function" [ACF1], residue correlation, expression (15):

$$ACF_k = \frac{cov(R_{it}, R_{i,t-k})}{variância(R_{it})} \tag{15}$$

### 2.7 Multivariate models with counting data

The test model base understood the canonical linkage of the efficiency variable of the counting type, by the likelihood method under Binary Logistic, Multinomial Logistic, Poisson, and Negative Binomial distributions. The multivariate analysis with counting data aimed to investigate efficiency under rare phenomena or with inflation of zeros. The dependent variable was transformed into "dummies": 1 was assigned to the efficient units and 0 to the others, allowing for the accumulation of frequency counts from 2015 to 2019, resulting in 5 observations and 6 possible frequency categories [0:5]. The analyzed sample was part of the Global set (60), without distinction of category or court function.

The explanatory variables were g7, tpesc, inf2, and varah (measured in units of justice per 100,000 inhabitants), incorporating the complete Global set (60) without distinction among court or justice categories. Seven models were configured: 1) logistic; 2) multinomial; 3) BNGE; 4) Bnge (s/g7 and s/varah); 5) Poisson; 6) ZINB; 7) ZIP. Table 2:

Table 2. Adjustments to the y_dummy variable [eff] for cumulative count data

| Optimization | Distribution | Categories | [eff_dummye] |
|---|---|---|---|
| 1 logistic | Bernoulli | 2 Cat | 1 = yes; 0 = no |
| 2 multinomial | Binomial | 3 Cat | (0-1); (2-3); (4-5) |
| 3 Bnge | Binomial | 2 Cat | 1 = eff 0 = no |
| 4 Bnge | Binomial | 6 Cat | cumulative [0.5] |
| 5 Poisson | Poisson | 6 Cat | cumulative [0.5] |
| 6 ZINB | Bernoulli Poisson-Gama | 6 Cat | cumulative [0.5] |
| 7 ZIP | Bernoulli Poisson | 6 Cat | cumulative [0.5] |

Source: Original survey results

The study was conducted in Excel®, utilizing the Solver function to maximize the maximum likelihood function, calculate the pseudo-R, and assess the significance of the variables using the X2 test (Fávero and Belfiore, 2022), under the following hypotheses:

Ho = ; H $\beta_1 = \beta_2 = \beta_3 = \beta_4 \ldots \beta_k = 0$ $_1$ = there is at least one ≠ 0.$\beta_j$

The statistic $X^2$ is calculated in expression (16):

$$X^2 = -2.(LLo - LLmax) \quad (16)$$

McFaden's pseudo-R2 is calculated in expression (17):

$pseudo$ R2 = (-2LLo – (-2LLmax))/-2LLo (17)

Where:

LLmax= is the maximum value of the likelihood function of the optimized model;

LLo = is the maximum value of the likelihood function of the null model.

### 2.7.1 Binary logistic regression model

The probability of occurrence of an event in the Bernoulli distribution is 1 for yes and zero for no. The probability function in binary logistic regression is given in expression (18):

$$\ln \frac{p_i}{1-p_i} = \alpha + \beta_1 . X_{1i} + \beta_2 . X_{2i} + \cdots \beta_k . X_{ki} \quad (18)$$

Where: p is the probability of the occurrence of a phenomenon of interest defined by Y=1, given that the dependent variable Y is a dummy variable.

The likelihood function is given in equation (19) and the maximization in (20):

$$L = \prod_{i=1}^{n} \left[ p_i^{Y_i} . (1 - p_i)^{1-Y_i} \right] \quad (19)$$

$$LL = \sum_{i=1}^{n} \left\{ \left[ (Y_i) . \ln \left( \frac{1}{1+e^{-Z_i}} \right) \right] + \left[ (1-Y_i) . \ln \left( \frac{1}{1+e^{Z_i}} \right) \right] \right\} = máx \quad (20)$$

### 2.7.2 Multinomial logistic regression model

The probability of occurrence of an event in the Binomial distribution, in equation (21):

$$\ln \frac{p_{im}}{1-p_{im}} = \alpha_m + \beta_{1m} . X_{1i} + \beta_{2m} . X_{2i} + \cdots \beta_{km} . X_{ki} \quad (21)$$

Where: $p_m$ (m = 0.1,... M-1) is the probability of occurrence of each of the M categories of the dependent variable Y.

The likelihood function is given in equation (22) and the maximization (23):

$$L = \prod_{i=1}^{n} \prod_{m=0}^{M-1} \left( p_{i_m} \right)^{Y_{im}} \quad (22)$$

$$LL = \sum_{i=1}^{n} \sum_{m=0}^{M-1} \left[ (Y_{im}) . ln \left( \frac{e^{Z_{i_m}}}{\sum_{m=0}^{M-1} e^{Z_{i_m}}} \right) \right] \quad (23)$$

### 2.7.3 Poisson regression model for counting data

The Poisson model is employed to analyze event counts, where the frequency of occurrences is distributed over a fixed interval of time or space. This model assumes that the mean and variance of the counts are equal, making it suitable for data with relatively uniform dispersion. The probability of an event's occurrence in the Poisson distribution is represented by equation (24): uniform dispersion. The probability of occurrence of an event in the Poisson distribution, equation (24):

$$\ln(Y_i) = \alpha + \beta_1.X_{1i} + \beta_2.X_{2i} + \cdots \beta_k.X_{ki} \quad (24)$$

Where: λ is the expected value of the quantity of occurrence of the phenomenon represented by the dependent variable Y, which presents count or *ratio data*.

The likelihood function is given in equation (25) and the maximization (26):

$$L = \prod_{i=1}^{n} \frac{e^{-\lambda_i}.(\lambda_i)^{Y_i}}{Y_i!} \quad (25) \qquad LL = \sum_{i=1}^{n} \left[ -\lambda_i + (Y_i).ln(\lambda_i) - ln(Y_i!) \right] \quad (26)$$

### 2.7.4 Negative binomial regression model for counting data

Models counts where the variance exceeds the mean (overdispersion), it allows the variance to surpass the mean, which is beneficial when there is additional variability in the counts. More flexible than the Poisson model for data with high variability, the probability of an event occurring in the Negative Binomial distribution (with the parameter for overdispersion, θ) is given (27):

$$\ln(\mu_i) = \alpha + \beta_1.X_{1i} + \beta_2.X_{2i} + \cdots \beta_k.X_{ki} \quad (27)$$

Where: μ is the expected value of the quantity of occurrence frequency of the phenomenon denoted Poisson_gama, Exponential Generalized Negative Binomial [BNGE].

The likelihood function of maximization is given in equation (28):

$$LL = \sum_{i=1}^{n} \left[ Y_i.\ln\left( \frac{\phi.\lambda_{bneg_i}}{1+\phi.\lambda_{bneg_i}} \right) - \frac{\ln\left(1+\phi.\lambda_{bneg_i}\right)}{\phi} + \ln\Gamma\left(Y_i+\phi^{-1}\right) - \ln\Gamma\left(Y_i+1\right) - \ln\Gamma\left(\phi^{-1}\right) \right] \quad (28)$$

### 2.7.5 Negative binomial regression model with inflation of zeros

It combines a negative binomial model for handling counts (with overdispersion) and a binomial model for addressing excess zeros. There are two parts: one models the probability of excessive zeros, and the other models the non-zero count using a negative binomial distribution. The model is suitable for data with many zeros and overdispersion. It is more flexible than other models for data with high variability. However, this model is more complex to implement and interpret. It requires more computational resources and a greater understanding of the data. The maximization of the likelihood function of the "Zero Inflated Negative Binomial" [ZINB] is given in equation (29), where θ represents the inverse of the parameter in the Gamma distribution:

$$LL = \sum_{Y_i=0} \ln\left[ p_{logit_i} + \left(1 - p_{logit_i}\right).\left(\frac{1}{1+\phi.\lambda_{bneg_i}}\right)^{\frac{1}{\phi}}\right] + \sum_{Y_i>0}\left[\ln\left(1-p_{logit_i}\right) + Y_i.\ln\left(\frac{\phi.\lambda_{bneg_i}}{1+\phi.\lambda_{bneg_i}}\right) - \frac{\ln\left(1+\phi.\lambda_{bneg_i}\right)}{\phi} + \ln\Gamma\left(Y_i+\phi^{-1}\right) - \ln\Gamma\left(Y_i+1\right) - \ln\Gamma\left(\phi^{-1}\right)\right] = \text{máx} \quad (29)$$

### 2.7.6 Poisson regression model and with inflation of zeros

The maximization of the likelihood function "Zero Inflated Poisson" [ZIP] is given in equation (30), where θ represents the inverse of the parameter. It is suitable for data with excess zeros and moderate variability.

$$LL = \sum_{Y_i=0} \ln\left[ p_{logit_i} + \left(1 - p_{logit_i}\right).e^{-\lambda}\right] + \sum_{Y_i>0}\left[\ln\left(1-p_{logit_i}\right) - \lambda_i + \left(Y_i\right).\ln\left(\lambda_i\right) - \ln\left(Y_i!\right)\right] = \text{máx} \quad (30)$$

## 3. Results and Discussion

### 3.1 Simple linear models

Table 3 analyzes the significance of 4 Global efficiency models (60), TJEstduals (27), TRTs(27), and TRFs(6) over a period of 7 years from 2014 to 2021. In the ANOVA analysis and the meaning of the F of the TJEstaduals, H0 was rejected at a significance level of less than 5%, falling below the three-digit mark. Therefore, there is at least one parameter of x other than 0. The correlation ρ of the TJEstaduals was positive and strong at 0.8877, indicating that efficiency is positively correlated with time in years.

Table 3. Test for parameter significance for simple linear regression

| Test Significance Parameters | Global (60) | TJEstaduals (27) | TRFs (6) | TRTs (27) |
|---|---|---|---|---|
| ANOVA F | 3,74778121 | 18,58997904 | 0,20264474 | 0,090070339 |
| F for Meaning | 0,11064789 | 0,007630983 | 0,67144076 | 0,776164694 |
| P-value intersection | 0,12237414 | 0,0082202 | 0,68892717 | 0,74164138 |
| P-valor variavel x | 0,11064789 | 0,007630983 | 0,67144076 | 0,776164694 |
| Standard error | 0,02551589 | 0,025148885 | 0,06730084 | 0,043446173 |
| Correlation (ρ) | 0,65454296 | 0,887719322 | 0,19735838 | -0,133023695 |

Source: original research results

Figure 3 illustrated the dispersion of the 7 points of the efficiency index over time, depicting the original regression line and the respective equations for the 4 simple linear models. According to Fávero and Belfiore (2022), the regression line with a zero intercept (y2) has the property of either excluding or maintaining the β parameter in the model. The regression line with an intercept of 0 for the TRFs and TRTs was included. The significance of the parameters led to the selection of the TJEstaduals set for the multivariate modeling stage.

Regarding the F of Signification of 0.7%, the efficiency of the TJEstaduals (27) was better explained in relation to time when compared to the TRFs(6), TRTs (27), or the Global model (60). Although the F of Meaning of the Global model (60) has provided a better explanation than the sets TRFs(6) or TRTs (27), there is evidence that the Global model (60) in the analysis of efficiency concerning time is influenced by the parameters associated with the set TJEstaduals (27).

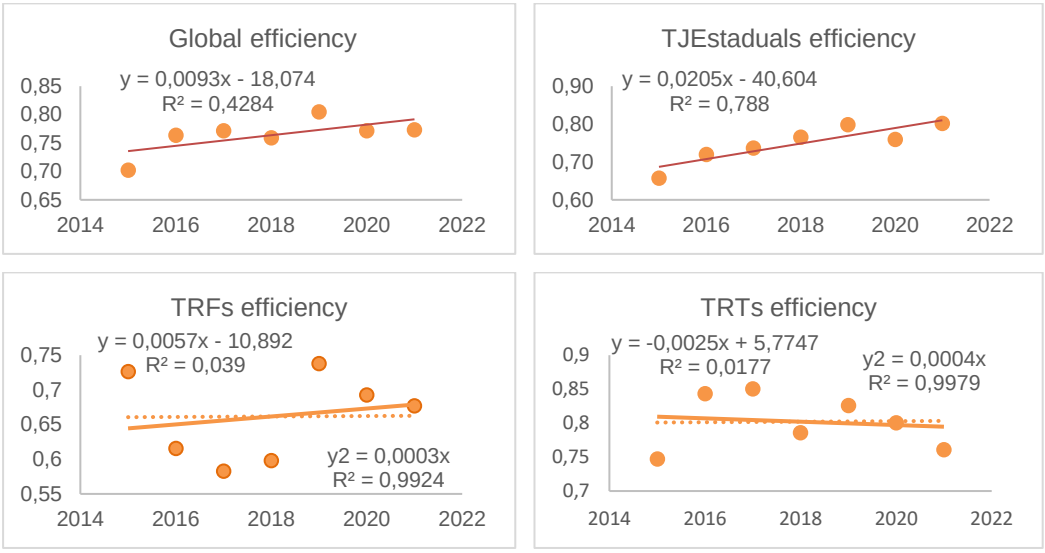


Figure 3. Original regression lines of efficiency with intercept equal to zero
Source: original research results

### 3.2 Multivariate GLMs of the TJEstaduals (27) from 2009 to 2022

The data series spans 14 years using RStudio (2024). The significance level of the Shapiro-Francia normality test was presented under the distribution of the dependent variable for the 27 units (Figure 4).

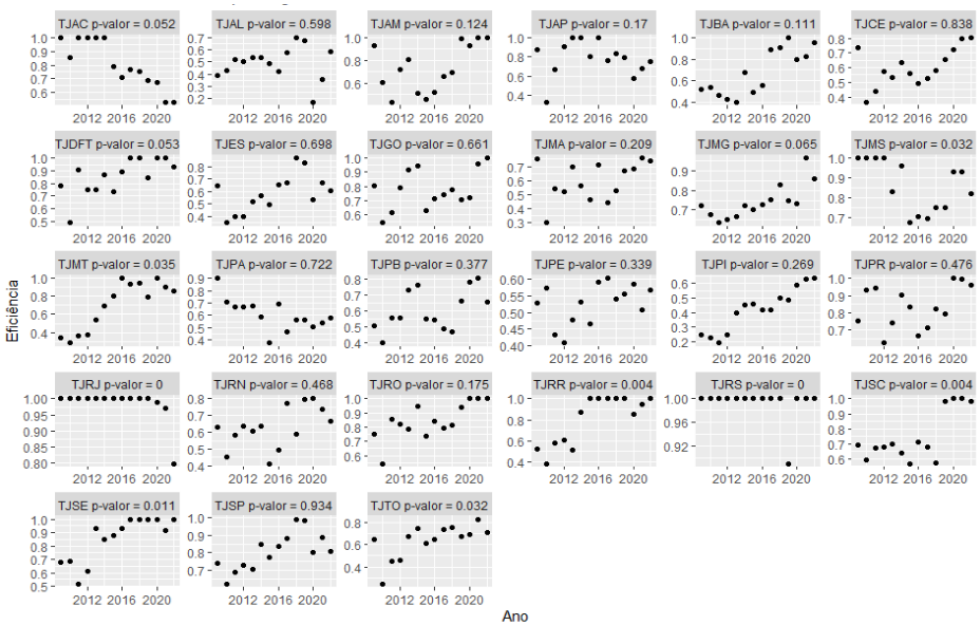


Figure 4. TJEstaduals eff and p-value normality Shapiro-Francia 2009- 2022
Source: original research results

From 2009 to 2022, the TJEstaduals had an average of 0.7223 efficiency, with an average of 8,159 new cases per year, with 1.05 computers per user, occupying 438,655 $m^2$, published an average of 68,865 second-degree judgments (case in the appellate phase) per year, with an average investment of R$ 246.64 reais per 100,000 inhabitants, with an average of 29 civil servants allocated to judicial and magistracy schools (Table 4).

Table 4. Exploratory statistical summary of explanatory variables

| Summary | eff | Ch | inf2 | Mtotal | apublic2 | g7 | TPESC | f5 | f6 |
|---|---|---|---|---|---|---|---|---|---|
| Min | 0,1697 | 2441 | 0 | 0 | 0 | 57,23 | 0 | 0,00361 | 0 |
| 1st Qu | 0,5677 | 5650 | 0,9349 | 160923 | 7931 | 151,1 | 12 | 0,01501 | 0,00257 |
| Median | 0,7179 | 7920 | 1,0327 | 298845 | 16971 | 210,1 | 24 | 0,0204 | 0,00612 |
| Average | 0,7223 | 8159 | 1,0581 | 438655 | 68865 | 246,64 | 29 | 0,0237 | 0,0052 |
| 3rd Qu | 0,9091 | 10529 | 1,1962 | 438655 | 49939 | 304,42 | 39 | 0,02841 | 0,00856 |
| Max | 1 | 17454 | 2,0275 | 543896 | 840073 | 1046,93 | 127 | 0,07427 | 0,02112 |

Source: original research results

A new visualization of data on maps (geoprocessing) was generated in QGIS 3.36.0 (2024), featuring the variables eff, ch, mtotal, and g7 displayed in a graded manner across quartiles based on the distribution in the Brazilian territory of the TJEstaduals from 2009 to 2022 (Figure 5).

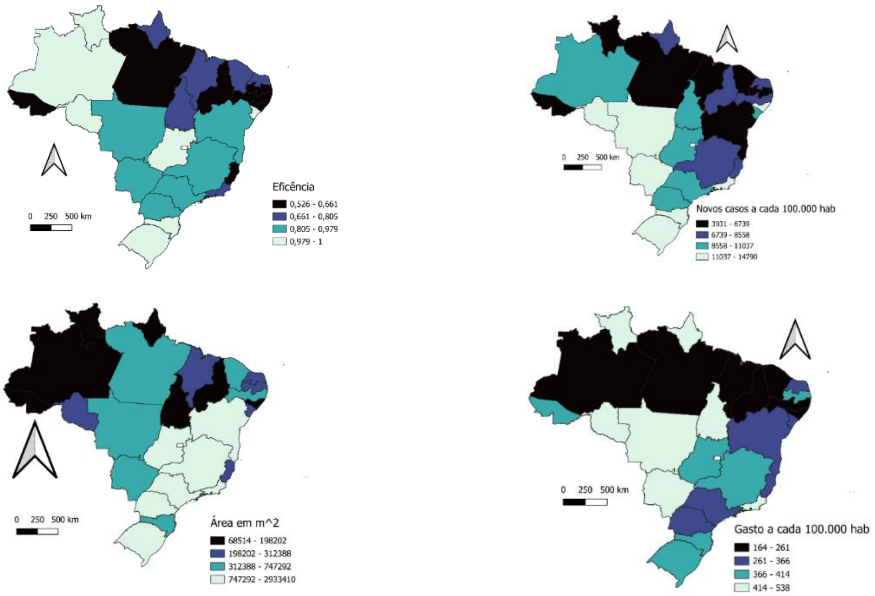


Figure 5. Data visualization State maps from eff, ch, Mtotal ($m^2$) and G7
Source: original survey results

In expression (30), the intercept and ch (new cases) were statistically significant at a level below 0.1%. The inf2 (computers per user) and f5 (proportional servers at [IT]) were significant at the 10% level. The variables tpesc and f6, related to learning, showed negative coefficients. The residual error of the model was 0.1289, with an $R^2$ of 53.62%. The missing variables are due to the history of tpesc, f5, and f6, whose records are from 2014 onwards (Table 5).

eff = intercept + $\beta_1$ch + $\beta_2$inf2 + $\beta_3$mtotal + $\beta_4$apublic2 + $\beta_5$g7 +$\beta_6$tpesc + $\beta_7$f5 + $\beta_8$f (30)

Table 5. Multivariate full model

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.26034 -0.10267 -0.01014  0.09018  0.30753
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.757e-01  5.368e-02   5.136 1.07e-06 ***
ch           2.722e-05  5.442e-06   5.002 1.91e-06 ***
inf2         9.671e-02  4.620e-02   2.093   0.0384 *
mtotal       4.925e-08  7.060e-08   0.698   0.4867
apublic2     1.614e-07  1.999e-07   0.807   0.4211
G7           1.275E-04  1.054E-04   1.210   0.2286
TPESC       -1.578E-05  6.148E-04  -0.026   0.9796
F5           2.628E+00  1.054E+00   2.493   0.0140 *
F6          -1.150E-01  3.193E+00  -0.036   0.9713
```

Source: original research results

The highest positive correlation of 0.85, with a significance level (ρ) of 0.01%, was observed between the variables mtotal and apublic2, indicating a greater geographic concentration for collegiate decisions in second instances. The second highest positive correlation occurred between eff efficiency and ch new cases. The inf2 did not show a correlation above 30% (Figure 6).

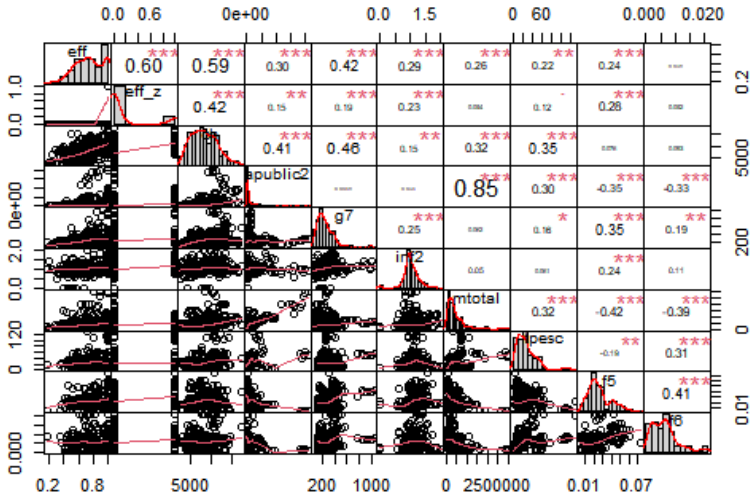


Figure 6. Correlation matrix with distribution and significance level
Source: original research results

To diagnose multicollinearity and heteroscedasticity, values above 10 suggest replacing dependent variable Y Efficiency with multicollinear variables (Fávero and Belfiore, 2022). The highest observed VIFs were 8.63 for mtotal and 7.84 for apublic2. The ch had a VIF of 2.16, with others falling below this, indicating low multicollinearity. The Breusch-Pagan test's low significance led to accepting H0, showing no heteroscedasticity effect; refer to Table 6 in the complete model.

Table 6. Diagnosis of multicollinearity and heteroscedasticity

Multicollinearity Diagnosis

```
> olsrr::ols_vif_tol(modelo_compl
  Variables Tolerance      VIF
1        ch 0.4623593 2.162820
2      inf2 0.8041936 1.243482
3    mtotal 0.1157651 8.638182
4  apublic2 0.1274911 7.843687
5        g7 0.5225731 1.913608
6     tpesc 0.5670471 1.763522
7        f5 0.5774792 1.731664
8        f6 0.5417537 1.845857
>       scale = F, digits =
```

Diagnosis Heteroscedasticity

```
-------------------------------
DF            =    1
Chi2          =    1.778904e-05
Prob > Chi2   =    0.9966348
scale = F, digits = 4)
```

Source: original research results

A total of five models for multivariate evaluation were studied (Figure 7): i) complete model; ii) Box-Cox model; iii) model excluding the variables Tesc, f5, and f6; iv) Step model iii; v) complete step model i. The step model (iii) also excluded the variable mtotal. The Box-Cox model did not demonstrate significant improvement and introduced a negative intercept. The full step model excluded the learning variables tpesc and f5.

| | Modelo completo | Modelo Box Cox | Modelo | Modelo Step | Modelo Completo Step |
|---|---|---|---|---|---|
| (Intercept) | 0.2757 *** | -0.6834 *** | 0.3014 *** | 0.3000 *** | 0.2752 *** |
| | (0.0537) | (0.0504) | (0.0422) | (0.0409) | (0.0523) |
| ch | 0.0000 *** | 0.0000 *** | 0.0000 *** | 0.0000 *** | 0.0000 *** |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| inf2 | 0.0967 * | 0.0930 * | 0.1188 *** | 0.1189 *** | 0.0964 * |
| | (0.0462) | (0.0434) | (0.0356) | (0.0356) | (0.0456) |
| mtotal | 0.0000 | 0.0000 | 0.0000 | | 0.0000 |
| | (0.0000) | (0.0000) | (0.0000) | | (0.0000) |
| apublic2 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| g7 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| tpesc | -0.0000 | 0.0000 | | | |
| | (0.0006) | (0.0006) | | | |
| f5 | 2.6283 * | 2.5073 * | | | 2.6276 ** |
| | (1.0543) | (0.9896) | | | (0.9848) |
| f6 | -0.1150 | -0.4047 | | | |
| | (3.1934) | (2.9974) | | | |
| N | 132 | 132 | 286 | 286 | 132 |
| R2 | 0.5362 | 0.5348 | 0.4025 | 0.4025 | 0.5361 |

*** p < 0.001; ** p < 0.01; * p < 0.05.

Figure 7. Summary of TJ Multiple Regression ModelsStates (27)
Source: original research results

### 3.3 TJEstaduals Time-series of the new cases ch from 2009 to 2022

There is a higher density in the 5,000 to 7,500 and 7,500 to 10,000 intervals, with an average of 8,159 new cases (Figure 8). ARIMA estimation can assume probability distributions from the exponential family, such as Poisson, logistic, or log-linear; these are classified as LGM (Fávero and Belfiore, 2022).

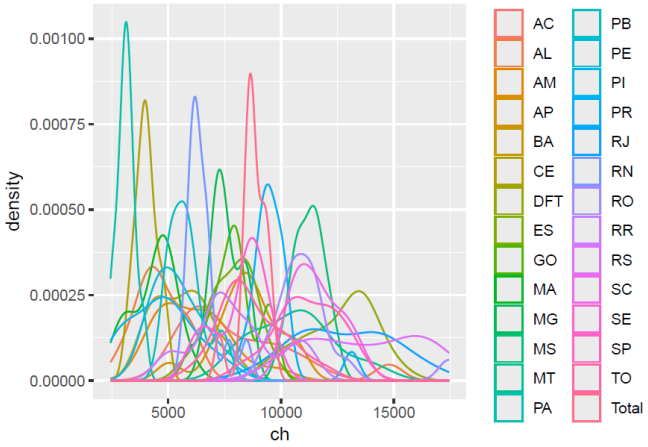


Figure 8. Density distribution of States of the relevant variable ch
Source: original research results

The analysis of the accumulated ch series was utilized for estimation in the ARIMA (0,0,0) model. The Dickey-Fuller stationarity test (1979) was statistically significant in accepting H0, meaning the original series is not stationary and is an integrated or differentiated process of order 1, as shown (Figure 9).

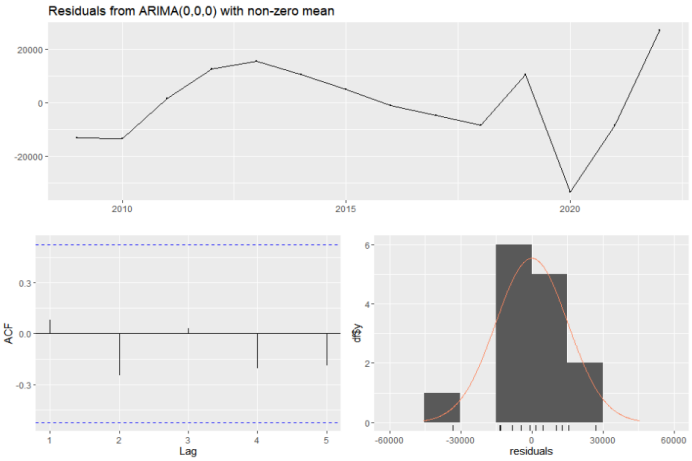


Figure 9. Cumulative series of TJEstaduals ch (new cases) for estimation
Source: original research results

Figure 10 displays the results of the stationarity test for three lags, using the best adjusted R optimized through the Ljung-Box residual test. Hossein and Yeganegi (2020) published an article on optimal lag selection via the Ljung-Box test, particularly when the value suggested by the automatic model is inappropriate for lag and the series remains non-stationary despite differentiations, especially in cases with few observations. They propose optimizing lag through the Ljung-Box test, which is sensitive to the number of lags used.

```
Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
   Min     1Q Median     3Q    Max
-24203 -11540  -3492  11463  26685

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
z.lag.1     -0.01329    0.02879  -0.462    0.661
z.diff.lag1 -0.86283    0.44620  -1.934    0.101
z.diff.lag2 -0.72252    0.48125  -1.501    0.184
z.diff.lag3  0.79523    0.79839   0.996    0.358

Residual standard error: 19800 on 6 degrees of freedom
Multiple R-squared:  0.4502,    Adjusted R-squared:  0.08363
F-statistic: 1.228 on 4 and 6 DF,  p-value: 0.3907


Value of test-statistic is: -0.4616

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.66 -1.95  -1.6
```

Figure 10. Unit root test result: the series is not stationary
Source: original research results

The Ljung-Box test does not indicate a significant correlation between the model's residuals nor does it detect the presence of the ARCH effect (Table 7).

Table 7. Ljung-Box test result and arch effects

```
> checkresiduals (estima_nc_t)

        Ljung-Box test

data:  Residuals from ARIMA(0,0,0) with non-zer
Q* = 1.2661, df = 3, p-value = 0.7372

Model df: 0.   Total lags used: 3
```

```
> ArchTest(estima_nc_t$residuals)

        ARCH LM-test; Null hypothesis: no ARCH eff

data:  estima_nc_t$residuals
Chi-squared = 2, df = 12, p-value = 0.9994
```

Source: original research results

Padulo and Ustari (2022) analyzed error statistics for forecasting time series models using ARIMA alongside other forecasting methods. The authors warned that the COVID-19 pandemic has significantly increased demand uncertainty due to changes in market behavior. The analysis for error minimization is optimized at lag 1 MAPE, considering the lower mean error shown in Table 8.

Table 8. Forecast statistics

| ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 | |
|---|---|---|---|---|---|---|---|
| 1.164153e-10 | 14654.29 | 11777.18 | -0.4241 | 5.2227 | 0.8692 | 0.0806 | (3) |
| -442.5069 | 13779.4 | 10298.24 | -0.5827 | 4.6313 | 0.7600 | 0.0317 | (2) |
| 105.3657 | 14586.3 | 11420.16 | -0.3751 | 5.06892 | 0.8428 | 0.0293 | (1) |

Fonte: original reserch results

The most optimal ARIMA model denoted as (0,0,0) with one lag. Figure 11 illustrates the subsequent two-year forecast. It is important to note that the forecast may be underestimated due to the implications of the COVID-19 pandemic.

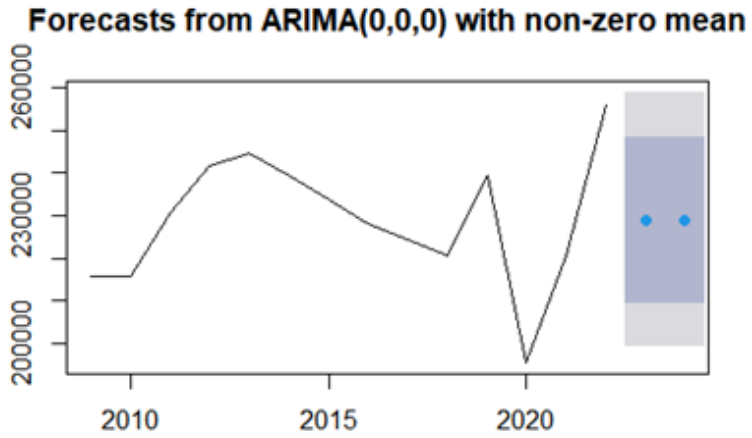


Figure 11. Best estimated model with next 2-year forecast
Source: original research results

Braga (2015) analyzed data from Public Institutions: Legislative, Judiciary, Executive, and Public Prosecutor's Office [PPO] to study the budget-expenditure relationship concerning initial credit. Evidence shows a connection between the Judiciary and PPO, but suggests studies with larger samples.

### 3.4 Global efficiency (60) for counting data from 2014 - 2022

In Table 9, seven counting models were compared based on their best adherence to the data. The selection criteria for the best model are based on optimizing the likelihood function [ optimized LL] or higher log-likelihood.

**Model 6 ZINB:** This model demonstrated the best adhesion and predictive performance, with an optimized likelihood value of -44.8. ZINB proved more suitable for the data due to its ability to manage many zeros and overdispersion, characteristics evident in the analyzed data. The pseudo-R was 0.98.

**Model 7 ZIP:** The Zero Inflated model achieved the second-best likelihood value of -62.19. Although it did not surpass ZINB, it still performed competitively, offering a robust alternative for modeling rare phenomenon.

**Model 4 BNGE:** The BNGE model achieved the third-best likelihood value of -66.2. Although it did not surpass ZINB, it still performed competitively, offering a robust alternative for modeling overdispersion. The difference between the **3Bnge** and the 4Bnge was the exclusion of g7 and varah in the 4Bnge, resulting in a higher likelihood from -129.2 to -66.2, with a pseudo-R of 0.73.

**Model 5 Poisson:** This model had a likelihood value of -83.7, ranking fourth. It was less effective than ZINB, ZIP, and BNGE, particularly in cases of excessive zeros and overdispersion. The pseudo-R was 1.

Table 9. Parameter Analysis of Likelihood Optimization

| Optimization | 1 Logistic | 2 Multinomia | 3 Bnge | 4 Bnge | 5 Poisson | 6 ZINB | 7 ZIP |
|---|---|---|---|---|---|---|---|
| LL initial | -194,0 | -4.396,6 | -1.426,7 | -249,9 | -57.020,9 | -2.908,0 | -190,2 |
| Optimized LL | -110,5 | -4.377,6 | -129,2 | -66,2 | -83,7 | -44,9 | -62,19 |
| Pseudo R | 0,43 | 0,00 | 0,91 | 0,73 | 1,00 | 0,98 | 0,67 |
| Test dist 5% | 166,97 | 37,99 | 2.595,09 | 367,28 | 11.3874,3 | 5.726,18 | 256,09 |
| Degrees Freedom (9.4877) | 4 | 4 | 4 | 2 | 4 | 4 | 4 |
| P-value | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 |

Source: original survey results

Table 10 presents the estimated coefficients for Model 6 of the Zero-Inflated Negative Binomial (ZINB) regression and those of other models. The analysis includes one intercept and four independent variables, detailed as follows:

- Intercept (α): -3.4545, representing the anticipated rate of events adjusted by the model parameters. A negative intercept indicates that, in the absence of other variables, the event rate is relatively low.
- The coefficient β1 (g7) is 0, signifying that this variable has no significant effect on the event rate when controlling for the study variables.
- Coefficient β2 (tpescola): 0.003, which indicates that an increase in this variable is associated with a small increase in the expected rate of events.
- Coefficient β3 (inf2): 0.4497, suggesting that the positive coefficient demonstrates a substantial influence of inf2 on the dependent variable.
- Coefficient of the variable β4 (varah): 0.0189, indicating that for each unit increase in varah, there is a slight increase in the event rate.

Table 10. Summary of Optimized Variables of the Likelihood Function

| Otimização | 1 logístico | 2 Multinomial | 3 Bnge | 4 Bnge | 5 Poisson | 6 ZINB | 7 ZIP |
|---|---|---|---|---|---|---|---|
| $\phi$ | x | x | 1,0000 | 2,3692 | x | 0,756 | x |
| $\theta$ | x | x | 1,0000 | 0,3721 | x | 1,323 | x |
| $\alpha$ | -6,320 | $\alpha 1$ e $\alpha 2$= 1 | 0 | 0 | 0,8333 | -3,4545 | -2,125 |
| $\beta_1$ | 0,0002 | $\beta 1$=1 | 0 | x | 0 | 0 | 0,028 |
| $\beta_2$ | 0,0221 | $\beta 12$= 0,6586 | 0,0020 | 0,106 | 0 | 0,003 | 0,336 |
| $\beta_3$ | 2,9098 | $\beta 22$ = 1 | 0,0940 | 0,001 | 0,011 | 0,4497 | 0,003 |
| $\beta_4$ | 0,1398 | x | 0 | x | 0 | 0,0189 | 0 |
| $\gamma$ | x | x | x | x | x | x | 2,096 |
| $\delta 1$ | x | x | x | x | x | x | -0,301 |

Source: original research results

The applicability analysis of counting models beyond the judicial system demonstrates their flexibility and robustness. They can be applied in Health for hospitalization rates (Poisson) and rare diseases with many zeros (ZINB); in Education for visits to websites and school absences (Negative Binomial); in Transportation for accidents or demand (ZINB); in Retail for the number of purchases or complaints (Poisson and ZINB); and in Marketing for interactions in digital campaigns (Negative Binomial and ZINB). Adapting to different sectors can optimize resources, predict events, and strengthen the study's relevance.

## 4. Conclusion

The performance efficiency of the judiciary system, connected to socioeconomic development, is aligned with a digital economy society, which includes pillars of digital transformation aimed at managing innovation as a trend in public policies for state reform. Sixteen generalized linear models were estimated:

- The initial study of four simple linear models related efficiency to time over seven years, from 2014 to 2021, using the sets of Global Justice units (60), TJEstaduals (27), TRTs (27), and TRFs (6). The significance of the sample parameters of the TJEstaduals set (27) led to the selection of multivariate modeling in RStudio®, incorporating eight variables from 2009 to 2022.
- TJEstaduals (27) units, among the study of five multivariate models, New Cases [ch], associated with digital transformation under the

pillars of entrepreneurship and society, was significant at $p<0.001$ in all models, with no multicollinearity or heteroscedasticity issues in the data. The variables inf2 and f5, related to the pillars of digital transformation in internal processes, were significant at $p<0.05$.

- Used QGIS® to visualize data of the quartile distribution of efficiency [eff], new cases [ch], expenditure per 100,000 inhabitants [g7], and area [Mtotal] across the 26 states and Federal District.
- The study of the principal component analysis of the time series of the relevant variable ch from the TJEstaduals set (27), using the Box-Jenkins method, produced an ARIMA (0,0,0) forecast with a lag of one, showing no ARCH effects or correlation of residuals. The projection over two periods indicated that it might be underestimated due to the effects of the COVID period.
- Among the seven models with count data from the Global set (60) units of the ecosystem from 2009 to 2022, the optimization of the likelihood function was identified in six ZINB models, explained by the variables g7, tpescola, the number of rods per 100,000 inhabitants [varah], and the relevant variable inf2.

The application of algorithms serves as a means of prediction and prevention, transforming parameters for the distribution of justice with supervenience. It aims to address the predominance of costs under the penal code, altering the constitutional order. Judiciary schools can lead knowledge management and lifelong learning, reflecting the role of legal institutions not only for control and safeguarding but, above all, for building bridges to socio-economic development. This approach shifts the critical mass of justice distribution from punishment and custody to prevention, information, and communication within the jurisdictional framework.

Science Analytics professionals with multidisciplinary training have the opportunity to integrate technological skills and soft skills, such as communication and collaboration, to enhance excellence in jurisdictional provision. Competitiveness must be analyzed in light of the principles of public administration, aiming to integrate the judicial ecosystem with economic and social development, where professionals and users are central to a more human-centered approach.

**Acknowledgments**

Thank you, God. I am grateful for my life. . I extend my thanks to the support of the MBA in Data Science and Analytics, Prof. Dr. Fávero, and his team. I thank Prof. Dr. Assaf Neto, coordinator of the MBA in Finance and Controllership, for my scholarship. I also appreciate the School of Economics, Administration, and Accounting at Ribeirão Preto (FEARP) for funding the publication. Lastly, I thank my family for their inspiring examples.

## 5. References

Akaike, H. 1974. A new look at the statistical model identification. in IEEE Transactions on Automatic Control. 19 (6): 716-723.

Almeida, V.A.; Almeida, V.A.; Araujo, F.P.O. 2021.Systematic mapping: methodologies and tools for teacher training in computational thinking. Journal New Technologies in Education [RENOTE]. 19 (2): 416–425.

Alon-Barkat, S.; Shamir, H. 2023. Israel's judicial overhaul and the role of the courts. *The Lancet*, 402(10397): 32–34. https://doi.org/10.1016/S0140-6736(23)01394-9

Alyas, T.; Abbas, Q.; Niazi, S.; Alqahtany, S.*; Alghamdi, T.; Alzahrani, A.; Tabassum, N,; Ibraim, A.* 2025. Multi blockchain architecture for judicial case management using smart contracts. *Sci Rep* **15**, 8471. https://doi.org/10.1038/s41598-025-92842-8

ANALYSISMACRO. 2024. ARIMA models and Box-Jenkins methodology. Available in ARIMA Models and Box-Jenkins Methodology - Macro Analysis (analisemacro.com.br). Accessed on: 2 Apr. 2024.

Beccaria, C. 1764. *Dei delitti e delle pene*.

Box, G.E.; Cox, D.R. 1964. An analysis of transformations. Journal of the Royal Statistical Society Series B: Statistical Methodology, 26 (2), 211-243.

Braga, P.L.F. 2015. Rate of budget increase of the Powers of the state of Minas Gerais: an analysis in temporal perspective and by data sampling. Graduate Program in Statistics of the Department of Statistics. Universidade Federal de Minas Gerais [UFMG], Belo Horizonte, Brazil.

Castro, M.V.B.; Balaniuk, R. 2020. DM-Trace: Traced data mining. Methodology for documenting data mining projects and its application in the construction of a textual classifier of documents associated with damage to the public purse. Journal of the Federal Court of Accounts. (145):39-79.

Chapman, P.; Clinton, J.; Kerber, R; Khabaza T.; Reinartz, T.; Shearer, C.; and Wirth, R. 2000. CRISP-DM 1.0. Step-by-step data mining guide.

Filgueiras, F., Mendonça, R. F., & Almeida, V. 2025. Artificial Intelligence and democracy: humans, machines and algorithmic institutions. *Estudos Avançados*, *39*(113), e39113075. https://doi.org/10.1590/s0103-4014.202539113.005

National Council of Justice [CNJ]. 2021. National Strategy Guide for Information and Communication Technology of the Judiciary [ENTIC-JUD] 2021-2026. Available at: compilado184145202110266178 4be9efedd.pdf (cnj.jus.br). Accessed on Apr. 07, 2024.

National Council of Justice [CNJ]. 2024. Justice in numbers database [DataJud]. Available at: Database - CNJ Portal. Accessed on: Apr. 05, 2024.

Dickey D.A.; Fuller, W.A.1979. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association ,74:427-431.

Djiroun, R.; Boukhalfa, K.; Alimazighi, Z. 2019. Designing data cubes in OLAP systems: a decision makers' requirements-based approach. Cluster Comput. 22: 783–803.

Emiliano, P.C. 2009. Fundamentals and applications of the information criteria: Akaike and Bayesiano. Master's thesis. Federal University of Lavras. Graduate Program in Statist and Agric Experiment. Lavras, Minas Gerais, Brazil.

Fávero, L.P; Belfiore, p. 2022. Data Analysis: statistics and multivariate modeling with excel, SPSS and Stata. LTC, Rio de Janeiro/RJ, Brazil.

Fávero, L.P.L.; Belfiore, P.P.; Silva, F.L.; Chan, B.L. 2009. Data analysis: multivariate modeling for decision-making. Rio de Janeiro: Elsevier.

Giancotti, M.; Rotundo, G.; Mauro, M. 2024. Factors affecting judicial system efficiency: a systematic mapping review with a focus on Italy. International Journal of Productivity and Performance Management. https://doi.org/10.1108/IJPPM-05-2023-0215.

Heine, F. 2017. Outlier Detection in Data Streams Using OLAP Cubes. In: Kirikova, M., et al. New Trends in Databases and Information Systems. ADBIS 2017. Communica in Computer and Information Science, Springer. 767:29-36.

Hess, H. C.2010. The principle of efficiency and the judiciary. R. Fac. Dir. Univ. SP.105: 211-239.

Hyndman, R.J.; Khandakar,Y. 2008. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 26 (3).

Kapopoulos, P.; Rizos, A. 2024. Judicial efficiency and economic growth: Evidence based on European Union data. Scottish Journal of Political Economy. 71: 101–131.

Lima, F.S; Marinho, E.L.L.; Costa, R.F.R. 2019. Non-Parametric Stochastic Production Frontier: an analysis of the Efficiency of the State Judiciary. 2019. Análise Econômica.37(72):159-186.

The National Audit Office UK [NAO].2025. Report Value for Money Reducing the backlog in the Crown Court. Available: Reducing the backlog in the Crown Court - NAO report. Accessed Apr. 2025

Nissi, E.; Giacalone, M.; Cusatelli, C. 2019. The Efficiency of the Italian Judicial System: A Two Stage Data Envelopment Analysis Approach. *Soc Indic Res.* 146: 395–407.

Organisation for Economic Co-operation and Development [OECD] .2024. Global Trends in Government Innovation 2024: Fostering Human-Centred Public Services, OECD Public Governance Reviews, OECD Publishing, Paris, https://doi.org/10.1787/c1bc19c3-en.

Paduloh, P.; Ustari, A. 2022. Analysis and comparing forecasting results using time series method to predict sales demand on covid-19 pandemic Era. JEMIS. Engineering & Management in Industrial System. 10: 37-49.

Pena, C. R. 2008. A model for evaluating the efficiency of public administration through the Data Envelopment Analysis (DEA) method. Journal of Administration and Accounting [RAC]. 12: 83-106.

QGIS 3.36.0. 2024. Free and open geographic information system. Available at: Welcome to the QGIS project. Accessed on April 14, 2024.

RStudio Desktop. 2024. Posit Software, PBC formerly RStudio, PBC. Available at: RStudio Desktop - Posit. Accessed on Feb. 05. of 2024.

Salazar-Salazar, G.; Mora, M.; Duran-Limon, H.A.; Rodríguez, F.J.Á. 2024. A Selective Comparative Review of CRISP-DM and TDSP Development Methodologies for Big Data Analytics Systems. In: Mora, M., Wang, F., Marx Gomez, J., Duran-Limon, H. (eds) Development Methodologies for Big Data Analytics Systems. Transactions on Computational Science and Computational Intelligence. Springer, Cham. Chapter:161-185.

Schiavon, L.C. 2021. Modernization of ways of working and performance management: an analysis of the digitalization of judicial processes in the productivity and efficiency of the Judiciary. National School of Public Administration [ENAP], Coleção Cátedras, Caderno 126, 58p, Brasília, Distrito Federal, Brazil.

Schröer, C.; Kruse, F.; Gómez, J.M.2021. A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Scienc. 181: 526-534.

Souza, M.M; Guimarães, T.A. 2018. Resources, innovation and performance in labor courts in Brazil. Journal of Public Administr. 52(3):486-506.

Stokel-Walker, C.; Van Noorden, R. 2023. What ChatGPT and generative AI mean for science. Nature. 614:214-216.

Tenório, T.R.S; Souza, F.G. 2021. What factors influence judicial efficiency? An analysis of the Brazilian State Courts of Justice. 21st USP Intern Conf in Account – Acc and Actua Sci improv eco and social devep, São Paulo.

Vasconcelos, M.A.S.; Alves, D.2000. Econometrics Manual. São Paulo, Atlas, Brazil.