

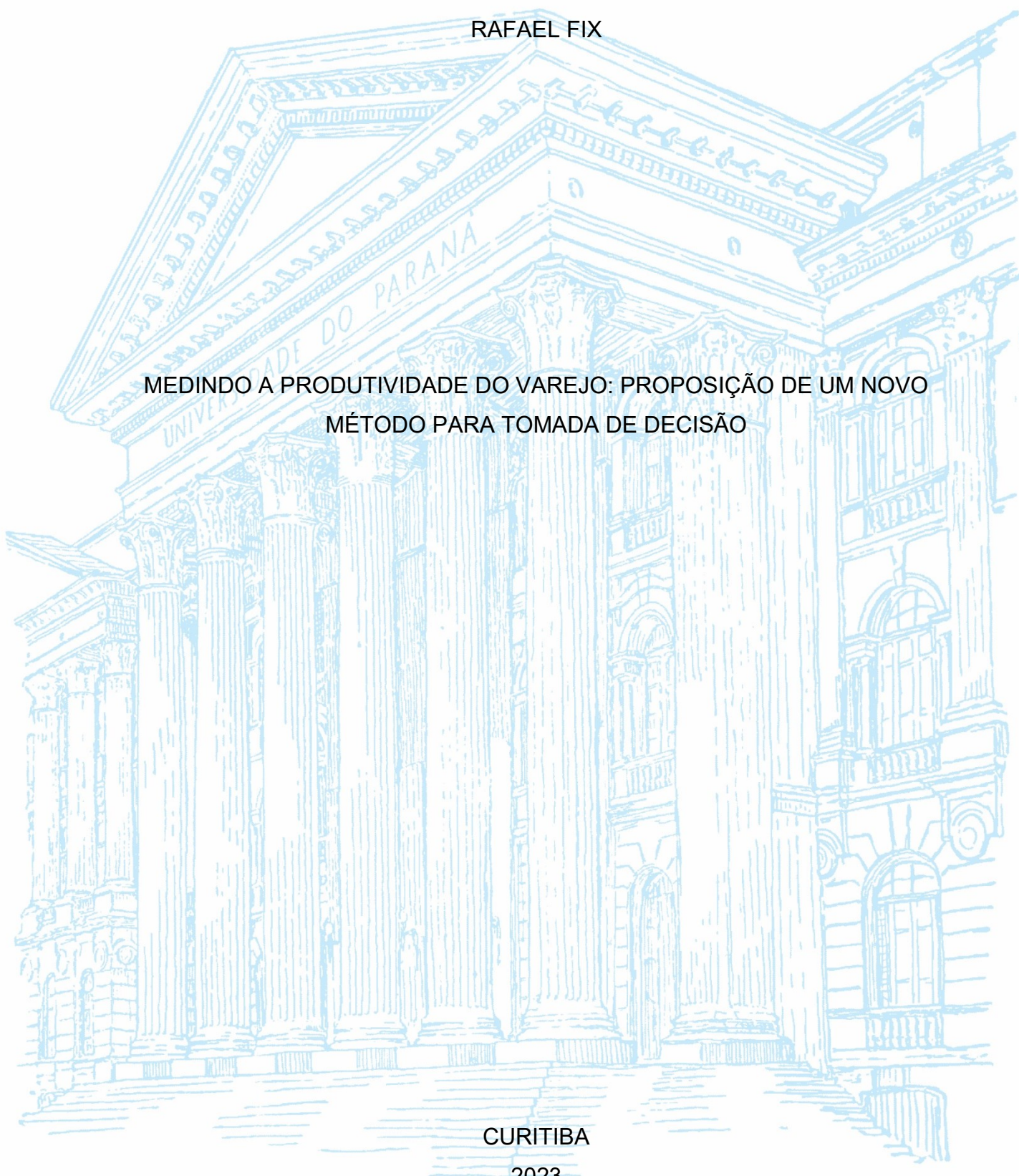
UNIVERSIDADE FEDERAL DO PARANÁ

RAFAEL FIX

MEDINDO A PRODUTIVIDADE DO VAREJO: PROPOSIÇÃO DE UM NOVO
MÉTODO PARA TOMADA DE DECISÃO

CURITIBA

2023



RAFAEL FIX

MEDINDO A PRODUTIVIDADE DO VAREJO: PROPOSIÇÃO DE UM NOVO
MÉTODO PARA TOMADA DE DECISÃO

Dissertação apresentada ao curso de Pós-graduação em Gestão de Organizações, Liderança e Decisão, Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Gestão de Organizações, Liderança e Decisão.

Orientador: Prof. Dr. José Eduardo Pécora Junior
Coorientador: Prof. Dr. Claudimar Pereira da Veiga

CURITIBA

2023

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIAS SOCIAIS APLICADAS

Fix, Rafael

Medindo a produtividade do varejo : proposição de um novo método para tomada de decisão / Rafael Fix. – Curitiba, 2023.
1 recurso on-line : PDF.

Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Ciências Sociais Aplicadas, Programa de Pós-Graduação em Gestão de Organizações, Liderança e Decisão.

Orientador: Prof. Dr. José Eduardo Pécora Junior.

Coorientador: Prof. Dr. Claudimar Pereira da Veiga.

1. Marketing - Canais de distribuição. 2. Comércio varejista - Produtividade. 3. Processamento de dados. 4. Análise de regressão. I. Pécora Junior, José Eduardo. II. Veiga, Claudimar Pereira da Veiga. III. Universidade Federal do Paraná. Programa de Pós-Graduação em Gestão de Organizações, Liderança e Decisão. IV. Título.

Bibliotecária: Maria Lidiane Herculano Graciosa CRB-9/2008



TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação GESTÃO DE ORGANIZAÇÕES, LIDERANÇA E DECISÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **RAFAEL FIX** intitulada: **MEDINDO A PRODUTIVIDADE DO VAREJO: PROPOSIÇÃO DE UM NOVO MÉTODO PARA TOMADA DE DECISÃO**, sob orientação do Prof. Dr. JOSÉ EDUARDO PÉCORA JUNIOR, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 20 de Setembro de 2023.

Assinatura Eletrônica

22/09/2023 17:45:52.0

JOSÉ EDUARDO PÉCORA JUNIOR

Presidente da Banca Examinadora

Assinatura Eletrônica

25/09/2023 14:19:21.0

TALITA MARIANA PINHO SCHIMIDT

Avaliador Externo (null)

Assinatura Eletrônica

22/09/2023 17:13:00.0

WESLEY VIEIRA DA SILVA

Avaliador Externo (UNIVERSIDADE FEDERAL DE ALAGOAS)

Dedico essa formação a minha esposa, que me inspira, incentiva e apoia em todas as minhas decisões; e aos meus filhos, que são a minha luz de cada manhã.

AGRADECIMENTOS

Agradeço ao meu orientador, Doutor José Eduardo Pécora Jr, cujo conhecimento e didáticas sempre admirei, desde que o conheci ainda durante a minha graduação. Pela sua compreensão e interesse em dividir meus objetivos.

Agradeço ao meu coorientador, Doutor Claudimar Pereira da Veiga, cuja notória inteligência e admirável carreira acadêmica e executiva viabilizaram um novo patamar para o meu mestrado.

Agradeço a todos os mestres e doutores que serviram de inspiração, que admiro e busco me equiparar, em especial os Doutores Marcos Wagner Fonseca, Felipe Fontes Rodrigues, José Roberto Frega, Humberto Dantas, Graziella Guiotti Testa Bruce e Carolina Fontanelli.

Agradeço a Sofia Berreta, Edward Sherwood-Smith e Luiz Gustavo Sanchez pelos aprendizados profissionais e incentivos.

RESUMO

Esta pesquisa propõe uma abordagem refinada para estimar a produtividade no setor de varejo utilizando um modelo matemático baseado na regressão linear multivariada *FGLS*. A inovação deste método reside na utilização da equação de erro gerada como um estimador para variáveis ocultas, que são empregadas para avaliar a eficiência de forma inferida. Essa metodologia foi aplicada em um conjunto de dados obtido de uma empresa líder no mercado de cereais prontos para consumo (*RTEC*), onde demonstrou resultados significativos. Em contraposição às tradicionais regressões de mínimos quadrados ordinários (MMQ), que normalmente oferecem resultados centrados na média, a metodologia proposta oferece informações de *benchmarking* relativas aos varejistas de desempenho superior, em oposição aos resultados típicos relativos à média das regressões. Além disso, diferentemente da análise convencional de envoltória de dados (DEA), esta nova abordagem utiliza a totalidade do conjunto de dados para definir o índice de eficiência. Isso supera as limitações inerentes à DEA, proporcionando maior liberdade e adaptabilidade na interpretação e análise dos resultados.

Palavras-chave: canais de marketing, produtividade no varejo, ciência de dados, mínimos quadrados generalizados viáveis.

ABSTRACT

This research proposes a refined approach to estimating productivity in the retail sector using a mathematical model based on FGLS multivariate linear regression. The innovation of this method lies in the use of the error equation generated as an estimator for hidden variables, which are used to assess efficiency in an inferred way. This methodology was applied to a data set obtained from a leading ready-to-eat cereal company (RTEC), where it showed significant results. In contrast to traditional Ordinary Least Squares (OLS) regressions, which usually provide results centered on the mean, the proposed methodology provides benchmarking information on top-performing retailers, as opposed to the typical mean regression results. Furthermore, unlike conventional data envelopment analysis (DEA), this new approach uses the entire data set to define the efficiency index. This overcomes the inherent limitations of DEA, providing greater freedom and adaptability in interpreting and analyzing the results.

Keywords: marketing channels, retail productivity, data science, feasible generalized least squares.

LISTA DE GRÁFICOS

| | |
|--|----|
| GRÁFICO 1 – HISTOGRAMA DE EFICIÊNCIA DOS VAREJISTAS | 26 |
| GRÁFICO 2 – ESTIMATIVA DA PRODUTIVIDADE DOS VAREJISTAS | 27 |

LISTA DE FIGURAS

| | |
|--|----|
| FIGURA 1 – CANAL DE MARKETING ATUAL E PRETENDIDO | 15 |
|--|----|

LISTA DE TABELAS

| | |
|--|----|
| TABELA 1 – CORREÇÃO DO ERRO | 22 |
| TABELA 2 – CORREÇÃO DA VARIÂNCIA..... | 22 |
| TABELA 3 - TRANSFORMAÇÃO E CORREÇÃO DOS DADOS..... | 25 |

LISTA DE ABREVIATURAS OU SIGLAS

- B2B2B - *Business-to-business-to-business*, empresa-para-empresa-para-empresa
- B2B2C - *Business-to-business-to-consumer*, empresa-para-empresa-para-consumidor
- DEA - *Data envelopment analysis*, Análise de envoltória de dados
- FGLS - *Feasible generalized least squares*, mínimos quadrados generalizados factíveis
- GLS - *Generalized least squares*, mínimos quadrados generalizados
- HFS - *High frequency store(s)*, varejista(s) de alto giro
- MMQ - Método dos mínimos quadrados
- OLS - *Ordinary least squares*, Método dos mínimos quadrados
- RTEC - *Ready-to-eat-cereals*, Cereais prontos para o consumo

LISTA DE SÍMBOLOS

© - copyright

@ - arroba

® - marca registrada

Σ - somatório de números

Π - produtório de números

SUMÁRIO

| | |
|----------------------------------|----|
| 1 INTRODUÇÃO | 15 |
| 1.1 PROBLEMA | 16 |
| 1.2 OBJETIVOS | 16 |
| 1.2.1 Objetivo geral | 16 |
| 1.2.2 Objetivos específicos..... | 16 |
| 1.3 JUSTIFICATIVA | 16 |
| 2 ARTIGO..... | 17 |
| REFERÊNCIAS..... | 30 |

1 INTRODUÇÃO

A transformação digital permite que as empresas alcancem maior flexibilidade e eficiência, otimizem processos produtivos, gerem propostas de valor para ecossistemas de inovação e respondam a tempo às necessidades do mercado (FELICIANO-CESTERO et al., 2023; GALINDO-MARTÍN et al., 2019). Então é bastante natural que uma grande indústria de cereais-prontos-para-consumo (RTEC, em inglês), como cereais matinais ou barras de cereal, tenha iniciativas nesta direção. A adoção de um modelo de ecommerce B2B2B (figura 1) é uma estratégia para aumentar o controle no canal de marketing (ALSAAD et al., 2021; BROWN & CROSNO, 2019). O modelo aqui chamado de B2B2B é um neologismo referenciando B2B2C, onde uma empresa intermedia a transação entre duas partes, um distribuidor/prestador a parte consumidora (G & ASOKAN-AKITHA, 2023), com a distinção de que a parte consumidora é, no caso, um varejista de alto giro (*HFS*, em inglês).

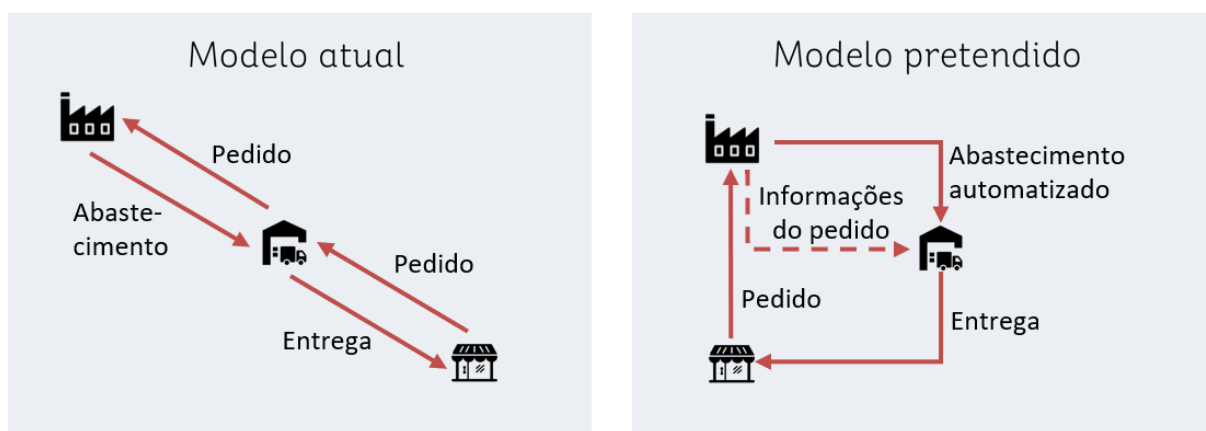


Figura 1: Canal de marketing atual e pretendido

Substituir um canal de alta persuasão por menos persuasivo (SHANKAR; KUSHWAHA 2021) precisa ter sua assimetria compensada por uma clara proposta de valor (EHRET; WIRTZ 2018). Pensando nisso, a indústria implementou um modelo de suporte à tomada de decisão que busca maximizar o sell-out (DING et al. 2019). O modelo consiste em estimar o maior volume de vendas possível, baseado em informações de outros varejos similares.

1.1 PROBLEMA

Para poder realizar recomendações precisas, é preciso uma forma de comparar os diferentes varejistas com diferentes características e restrições. Para isso é necessária uma forma de definir quanto cada varejista pode vender. Esta estimativa ainda não possui métodos consolidados e amplamente aceitos como ideais

1.2 OBJETIVOS

Este artigo propõe um método para inferir a produtividade de uma loja de varejo por meio da criação de uma equação com base em características disponíveis e estimando a máxima produtividade que cada loja de varejo pode alcançar.

1.2.1 Objetivo geral

Propor um modelo matemático que estime a eficiência atual e capacidade de faturamento do varejista (considerando sua capacidade a 100% da eficiência), para produtos da indústria, utilizando como parâmetros de entrada informações disponíveis em base de dados.

1.2.2 Objetivos específicos

- a) Determinar equação que estime a capacidade do varejo;
- b) Determinar equação que estime a eficiência do varejo.

1.3 JUSTIFICATIVA

O estado da arte utiliza DEA como principal ferramenta de estimativa (COOPER et al. 2011; DALVAND et al. 2014 e DONTHU; YOO 1998). E embora o modelo matemático seja bastante coerente, existem algumas críticas que inviabilizaram seu uso. O DEA utiliza somente os dados de vanguarda para fazer a estimação (DONTHU; YOO 1998), ignorando os demais dados. O DEA também assume que picos de eficiência podem ser mantidos por longos períodos, ou extrapolados sem nenhuma restrição (CABALLERO; RAHMAN 2014).

Measuring retail productivity: proposition of a new method for decision making

Rafael Fix¹

Jose Eduardo Pécora Jr¹

Claudimar Pereira da Veiga^{1,2}

¹ Graduate Program in Organizations Management, Leadership and Decision (PPGOLD), School of Management, Federal University of Parana (UFPR); Curitiba 80210-170, PR, Brazil

² Fundação Dom Cabral (FDC); Av. Princesa Diana, 760 Alphaville, Lagoa dos Ingleses, Nova Lima 34018-006, MG, Brazil

Abstract

This paper proposes an elegant way of estimating retail productivity through a mathematical model inspired by FGLS multivariate linear regression, using the error equation obtained as a hidden variable estimator to be used as inferred efficiency. This methodology was applied to a dataset obtained from a leading ready-to-eat cereal (RTEC) manufacturer and demonstrated substantial results. Unlike conventional ordinary least squares (OLS) regressions, the proposed approach offers benchmarking information relative to the best-performing retailers, as opposed to the typical relative-to-average outcomes of regressions. Furthermore, in contrast to traditional data envelopment analysis (DEA), this methodology relies on the entire dataset to establish the efficiency index, thereby addressing limitations imposed by DEA and allowing for executive discretion and flexibility in the analysis.

Keywords: Marketing Channels, Retail productivity, Data Science, Feasible Generalized Least Squares

INTRODUCTION

Analyzing productivity in retail is imperative for companies, being a challenge, mainly in developing predictive models that can help managers in decision-making (Dubelaar et al., 2002). Retail productivity has not - to this day - a consolidated approach, and it may not have since it is a very abstract estimation (Higón et al. 2010). It is widely focused on the Labor Productivity of the service provided (Ark et al., 2003; Doms et al., 2004; Haskel; Sadun, 2012; Lewis et al., 1998; O'Mahony; De Boer, 2002), hence service being considered a labor-intense activity. Other approaches consider a complete function of factors (inputs) that produce different outputs evaluated (revenue, margin, capacity factors) as performance indicators. However, even in these cases, models that consider a function still rely on data envelopment analysis (DEA) to make

estimates. DEA brings many advantages, such as accepting multiple outputs and directly providing an index that can be used as performance indicator, but it also comes with limitations. One of the main points to consider is that when constructing the efficiency frontier, all data points that do not lie on the frontier are not taken into account to calculate the performance index. This hinders the identification of outliers and affects the reliability of the metric. In this context, the DEA has limitations because it does not have the predictive capacity, requiring the development of new models (Dalvand et al. 2014, Shin et al. 2022).

This article proposes a method to infer the productivity of a retail store by creating an equation of its outcome based on available characteristics and estimating the maximum productivity each retail store can achieve.

A ready-to-eat-cereal (RTEC) industry channels its products to brick-and-mortar HFS (high-frequency stores, such as convenience stores, Mom and Pops, and gas stations) through exclusive distributors in each state. With the goal of digitizing this marketing channel and having greater management control, the industry is introducing a B2B2C digital channel, where retailers access the manufacturer's e-commerce platform, place their orders, and are supplied by the distributor. As a way to promote the channel, the industry has adopted the strategy of offering commercial intelligence, and for that, it seeks to build its recommendations based on the retailer's in-stalled capacity, not its current consumption.

Web channels are found to be less persuasive (Shankar; Kushwaha 2021), and the decision of how to maximize efficiency while considering performance and value generation is one of the main challenges faced by the industrial marketing manager (Coughlan 2010). On the other hand, it is important for retailers to assess the contribution of the industry to their results (Ailawadi; Farris 2017).

The proposed challenge is to increase the perceived value of the digital channel for the retailer by using product mix recommendations and commercial intelligence information about similar-sized retailers as a persuasion tool (Haskel; Sadun 2012), helping the retailer increase downstream

sales and consequently improve operational results. The constraint imposed is that the sell-in must be equal to the sell-out, avoiding inventory and operational cost increases, or margin reductions (Ailawadi; Farris 2017 e Hutt; Speh 2016). To achieve this goal, it is essential to accurately estimate the commercial capacity of each retailer (maximum installed productivity).

MATERIALS AND METHODS

2.1 Analytical Approach

2.1.1 Retail Productivity

Since Herbert Heaton (1977), there has been an ambiguity in the definition of productivity, efficiency, and effectiveness. Achabal et al. (1984) consolidate the econometric approach to this definition, where productivity relates to the installed capacity in a function of labor and capital. Efficiency is how this capacity is used, and effectiveness is the long-term indicator of the strategic decisions behind the efficiency installed.

Translating this definition to a practical example, productivity would relate to the size of the staff, inventory, and physical structure; efficiency is the return on investment (ROI) of the revenue over the costs of maintaining the given productivity. Therefore, a luxury brand would strategically choose a considerable productivity with lower efficiency since service level is key, and time would tell if that is an effective decision.

A relevant highlight for this definition is that sales should not be directly used as a metric for productivity because there might be outside factors playing a role (demand, premium pricing strategies) (Heaton, 1977), but this can be relieved, particularly if the comparison is among the same product mix with fixed price, as is the object of this paper (Lusch; Serpkenci 1990).

Donthu and Yoo (1998) bring a table with commonly used variables at the productivity function. As output, revenue, margins, and customer perception (service quality, satisfaction) and as inputs, environmental, customers, managerial and personal criteria.

2.1.2 Retail Productivity Estimation

This paper proposes a method derived from linear regression, which is not explored due to the nature of centering the prediction on the average performance. But as highlighted by Donthu and Yoo (1998), it is relevant to perform productivity estimates on a benchmark approach, comparing to top performers. This is naturally achieved using DEA, and that is why the majority of the research uses this method. DEA is designed to address many of the difficulties in productivity measurement. It allows for multiple input and output variables, produces an index that can be interpreted as a productivity factor, creates a frontier curve, and does not rely on a linear equation.

2.1.3 Heteroscedasticity.

Heteroscedasticity is defined as a variance that depends on the values of the exogenous variables (Colonescu 2016), as shown in Equation 1:

$$v(X) = Var(u | X)$$

This econometric phenomenon is organic and commonly observed in time series data, and is supported by theory in the analyzed data model: it is reasonable to expect that two retailers with the same characteristics (such as area and foot traffic) may have different sales figures and that an increase in these characteristics (such as larger store areas) may lead to an increase in the variation among retailers (Fomby et al. 1984).

2.1.4 Hidden variable.

Heteroscedasticity can also be explained by the presence of a hidden variable. For example, the relationship between price and property area may exhibit heteroscedasticity, but introducing additional variables, such as location, year of construction, and quality of finishing, may help to minimize this variance. The presence of unmapped variables – and hence hidden variables – can account for the variance in the error (Downs; Rocke 1979).

2.1.5 Regression Model

The FGLS (feasible generalized least squares) model is a multivariate linear regression model that aims to minimize the impact of heteroscedasticity compared to the Ordinary Least Squares (OLS) method (White; White1 1980), as shown in Equation 2.

$$\beta_{FGLS} = (X' \hat{\Omega}^{-1} X)^{-1} (X' \hat{\Omega}^{-1} Y)$$

WLS (weighted least squares) is another method for minimizing variance, which requires knowledge of the form (cause) of heteroscedasticity, while FGLS is a generalized form of regression. However, a kernel must be assumed to generalize the estimate (Bai et al. 1977). Tables 1 and 2 present the most commonly used corrections for error and variance (Miller; Startz 2018)

| Table 1. Error correction |
|----------------------------------|
| Residual transformation |
| \hat{U} (no transformation) |
| \hat{U}^2 |
| $\ln \hat{U}^2$ |
| $\ln \hat{U} $ |

| Table 2. Variance correction | |
|-------------------------------------|------------------------|
| $Var(u \bar{X})$ | Correction |
| Homoscedasticity | σ^2 |
| Moderate Heterosc. | $\sigma^2(\bar{X})$ |
| Severe Heterosc. | $\sigma^2 e^{\bar{X}}$ |

In this model, the quadratic residual transformation is used (as presented in equation 6), which has the effect of amplifying the residual, and the variance correction to account for severe

heteroscedasticity (as presented in equation 7), which prevents the existence of a negative bias.

The estimated FGLS regression model is presented in Equation 3.

$$\frac{Y_i}{\hat{h}(X_i)} = \frac{X\beta}{\hat{h}(X_i)} + \frac{u}{\hat{h}(X_i)}$$

2.2 Mathematical Model

The mathematical model is divided into three stages: (i) obtaining the error function (equations 4-8), (ii) defining a productivity estimator (equations 9,10), and (iii) predicting the productivity of retailers (equations 11, 12).

(i) Obtaining the error equation. To obtain the error equation, we partially applied the FGLS model in a linear regression model, where X is a matrix of estimators (Table 3). The first step is to perform an OLS regression on the equation:

$$Y = X\beta + u$$

To obtain the residual \hat{U} after the OLS regression of equation (4) by subtracting from the original estimators:

$$\hat{U} = Y - X\hat{\beta}_{OLS}$$

To define a second vector \hat{Y} , the square of the regression residual (5), according to the transformation proposed in table 1:

$$\hat{Y} = (\hat{U})^2$$

With this transformation, a second OLS regression is performed:

$$\hat{Y} = X\beta + e$$

And with the residuals $\hat{Y}_{residuals}$ obtained in Equation 7, the transformation for severe heteroscedasticity correction is applied - as specified in Table 2, defining $\hat{h}(X)$ equation as the exponential function of the vector $\hat{Y}_{residuals}$, as the equation for the error variance as a function of X :

$$\hat{h}(X) = e^{\hat{Y}_{residuals}}$$

The next step in an FGLS regression would be to apply error correction to the linear equation, as described in equation 3, and finally perform a new OLS regression on this equation – which is not performed in this method.

(ii) Productivity Estimation. In the proposed method, this is the point at which the leverage with the FGLS method cease. Once the error equation $\hat{h}(X)$ is obtained, the estimation equation for retail productivity can be constructed at equations (9, 10).

$$X_{productivity} = \hat{h}(X)$$

And included in the model:

$$Y_{corrected} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \beta_{productivity} x_{productivity}$$

(iii) Productivity projection. Finally, equations 11,12 show the final stage to determine the productivity, by first defining the estimator for productivity to 100%.p

$$x_{productivity} = 1$$

In the equation model (12), the projection of the retail performance is obtained.

$$Y_{predicted} = X\beta + \beta_{productivity} \times 1$$

This is the demonstration of the steps required to construct the model proposed in this article.

2.2.1 Application

A dataset was obtained from a major RTEC industry containing retailers that sell their products, channeled through exclusive state-defined distributors. The data reflects the fiscal year of 2021, includes several characteristics of each retailer, including the average weekly purchase amount, totaling 181 columns and 18,675 rows. The data science methodology described in the IBM white paper was followed, particularly, specifying the analytical approach, data correction, and modeling (Rollins, 2015). After identifying the relevant information and removing data that could not be used in the mathematical model (e.g., address, codes, unidentified information), 13 columns were kept being used in the model. Information such as geographic coordinates, area code and postal addressing code were removed, as although they could have a correlation with the prediction, they did not have a theoretical justification to be included in the model. Other location-related information, such as the average income of the population, proximity to points of interest (e.g. pharmacy, schools), was kept in the model but later discarded due to poor data quality. Table 3 presents the description of incomplete or distorted data.

Table 3. This is a table. Tables should be placed in the main text near to the first time they are cited.

| Variable | Missing information | Correction / transformation |
|---|---|--|
| Revenue (of products from this manufacturer) | 21 NaN (0.1%) | Rows dropped |
| Average income at ZIPCODE | 226 NaN (1.2%) | Rows dropped |
| Part of a chain? | 0 (0%) | Transformed do binary |
| Size code (Nielsen) | 0 (0%) | Transformed to dummy |
| Square footage | 0 (0%) | n/a |
| Ethnicity (% of pop. White, African, Hispanic, and Asian) | 178 NaN (1%) 12653 not adding up to 100% (67.7%) | Rows dropped Prorated to 100%, using Lagrangian relaxation ¹ |
| At/by a gas station? | 34 NaN (0.2%) | Transformed to fuzzy Replaced by average (0.886) |
| TOTAL | | 13 columns × 18,439 rows |

($\%_{ethnicities} + 0,0000001$) / $\sum_i (\%_{ethnicity_i} + 0,0000001)$, being i the number of ethnicities.

RESULTS

By applying the method to the dataset, it was possible to obtain an error function that, as per equation 9, reflects the productivity of each store. By following the proposed model Figure 1 contains a histogram of the productivity vector $X_{\text{productivity}}$ obtained at equation 9. Each bar has a width representing a range of 1% of productivity. It is possible to observe that the majority of retailers have a usage of its productivity between 20 and 40%, with the mode at 27 to 28% range. Very few retailers surpass the 75% threshold. Figure 2 displays 350 randomly selected retailers out of 18,439 available in the dataset. They are completely random, to allow for a better visualization of the result. The gray bar represents the average weekly revenue of each store – which is the available output factor used in this analysis, while the red line is the equation 12, projecting maximum productivity possible of being obtained for the given retailers, based on the parameters available. The entries are sorted ascending according to the estimated productivity capacity. Some information can be observed from the graphic, such as apparent discretization and three steps for the retailers with lower sales volume. This reflects the same discretization in the outcome variable. In the region of retailers with higher capacity, the curve becomes more sensitive. Another interesting feature is the distance between capacity and average sales, which is more pronounced for low-performing retailers than for those with higher productivity.

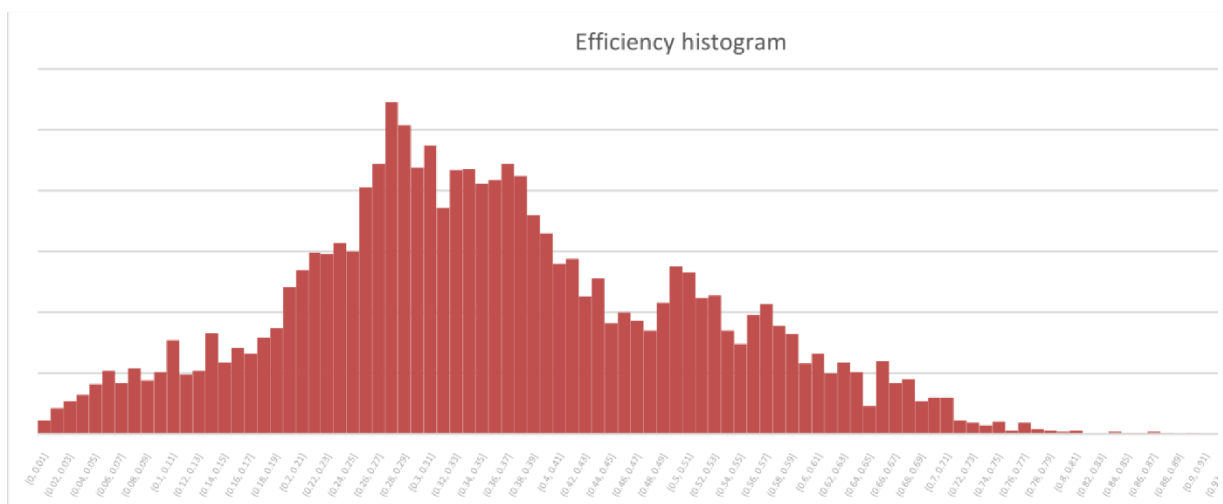


Figure 1. The histogram has all the efficiencies of 18.439 retailers. It is possible to observe the mode in the range of 27% to 28% of efficiency, while very few retailers have an efficiency above 75%.

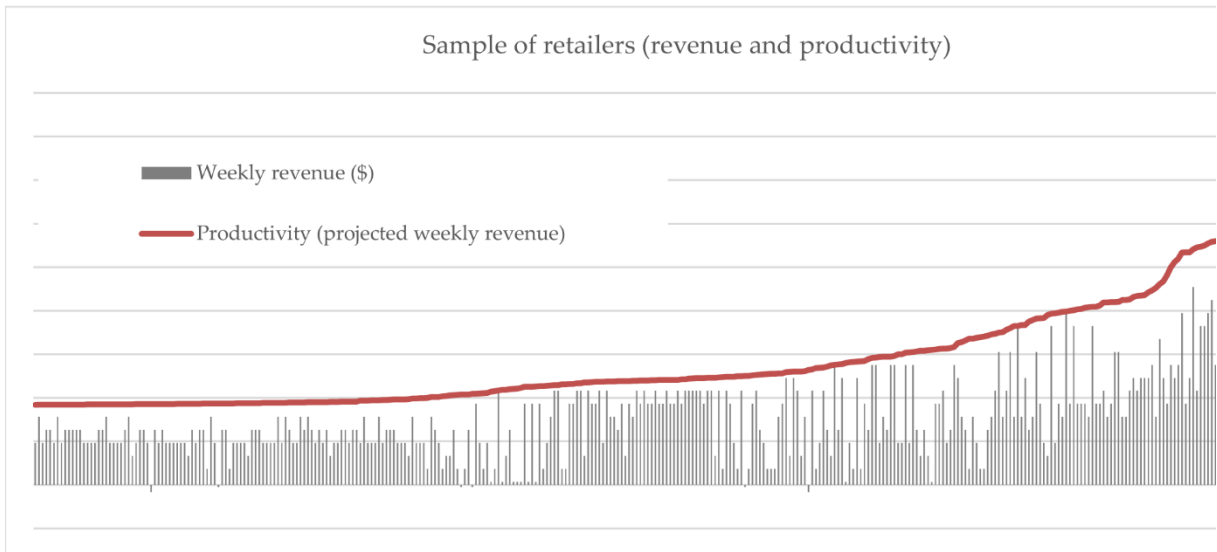


Figure 2. The figure contains a random sample of 350 retailers out of the dataset of 18.439 for better visualization. Each retailer is represented by a vertical bar of its weekly revenue and sorted by projected productivity ascendent (red line).

DISCUSSION

The objective of this model is not to correct for heteroscedasticity in the data, and therefore no heteroscedasticity test is performed. Rather, the model is inspired by extensive mathematical modeling available in the literature. The literature on retail performance questions the use of linear regressions for productivity models for considering as a limitation the tendency towards centrality (average productivity). This model understands that unlike the approach proposed by the literature using DEA, the information from the average is more robust than the efficiency frontier. However, this does not eliminate the need to understand the said frontier. The error equation indicates the individual distances between the retailer and the segment's average, with a sum of zero. These errors should follow a normal distribution, indicating that the transformations adopted in the residuals were appropriate. Thus, it is possible to convert the errors into an index ranging from 0 to 1 and consider this value as the retail efficiency. There are still discretionary decisions that need to be made, such as how to interpret the results and room for executive expertise. But it is definitely a model that can provide more information than a simple DEA efficiency index.

Although the number of predictive variables available in the dataset was limited, it was possible to verify that the mathematical model is capable of inferring productivity and the efficiency. This model still needs improvement, particularly on the as-assumptions of the FGLS method, determining which corrections have better adherence, and applying the model using the support vector machine artificial intelligence model proposed by Miller and Startz (2018). There is also the caveat that this analysis predicts internal capacities, but it may be subject to external limiting factors (competition, demand) (Downs; Roche (1979). A numerical comparison with DEA productivity analysis should also take place.

FINALS CONSIDERATIONS

This article contributes to the literature on productivity measurement by proposing a method to infer retail productivity based on available characteristics and estimate each store's maximum productivity. The model was tested on a composite data set of measures for productivity at a leading manufacturer of ready-to-eat cereals.

Given the limitations in making predictions using the DEA method to measure retailer productivity better, the proposed model was inspired by extensive mathematical modeling available in the literature, such as those presented in equations 1-10. As a result, the proposed model (equations 11 and 12) and the results shown in Figure 2 can provide more information than a simple DEA efficiency index, capable of inferring productivity and efficiency in retail by bringing empirical evidence of a practical case.

This article sheds light on the literature by demonstrating that retail productivity can differ if different strategies are used to measure model performance. Furthermore, the results indicate that it is possible to identify the best algorithm associated with better performance. Faced with market competitiveness, an approach to improve retail productivity can be helpful for companies by incorporating a model into a business intelligence system.

The main limitation of this study is related to the restriction of the model and analysis. Proposing a new model in the literature is a complex topic, considering the need for further tests for empirical validation since this study was not intended to exhaust the theme. Therefore, new studies are suggested, mainly for works with practical applications to test new parametric and non-parametric models that can help managers in decision-making in the most diverse retail segments. Future studies are suggested to test new performance models and constructs that combine parametric and non-parametric modeling.

REFERÊNCIAS

ACHABAL, D.; HEINEKE, J. M.; MCINTYRE, S. H. Issues and Perspectives on Retail Productivity - Achabal - Journal of Retail fall 1984. **Issues and perspectives on retail productivity**, v. 60, n. 3, p. 107–129, 1984.

AILAWADI, K. L.; FARRIS, P. W. Managing Multi- and Omni-Channel Distribution: Metrics and Research Directions. **Journal of Retailing**, v. 93, n. 1, p. 120–135, 2017. Elsevier Ltd.

ALSAAD, A.; TAAMNEH, A.; SILA, I.; ELREHAIL, H. Understanding the global diffusion of B2B E-commerce (B2B EC): An integrated model. **Journal of Information Technology**, v. 36, n. 3, p. 258–274, 2021. SAGE Publications Ltd.

ARK, B. VAN; INKLAAR, R.; MCGUCKIN, R. H. “Changing Gear”: Productivity, ICT and Service Industries in Europe and the United States. **The Industrial Dynamics of the New Digital Economy**. p.56–100, 2003. Edward Elgar Publishing.

BAI, J.; SUNG, ·; CHOI, H.; LIAO, · YUAN. Feasible generalized least squares for panel data with cross-sectional and serial correlations. **Empirical Economics**, v. 60, p. 309–326, 1977. Disponível em: <<https://doi.org/10.1007/s00181-020-01977-2>>. .

BALBONI, B.; TERHO, H. Outward-looking and future-oriented customer value potential management: The sales force value appropriation role. **Industrial Marketing Management**, v. 53, p. 181–193, 2016. Elsevier Inc.

BROWN, J. R.; CROSNO, J. L. Process and output control in marketing channels: toward understanding their heterogeneous effects. **Journal of Business and Industrial Marketing**, v. 34, n. 4, p. 735–753, 2019. Emerald Group Holdings Ltd.

CABALLERO, W. L.; RAHMAN, A. Application of Monte Carlo simulation technique for flood estimation for two catchments in New South Wales, Australia. **Natural Hazards**, v. 74, n. 3, p. 1475–1488, 2014. Kluwer Academic Publishers.

CLARKE, A. H.; FREYTAG, P. V. An intra- and inter-organisational perspective on industrial segmentation: A segmentation classification framework. **European Journal of Marketing**, v. 42, n. 9–10, p. 1023–1038, 2008.

COLONESCU, C. **Principles of Econometrics with R**. 2016.

COOPER, W. W.; SEIFORD, L. M.; ZHU, J. Data envelopment analysis: History, models, and interpretations. **International Series in Operations Research and**

Management Science, v. 164, p. 1–39, 2011. Springer New York LLC. Disponível em: <https://link.springer.com/chapter/10.1007/978-1-4419-6151-8_1>. Acesso em: 26/2/2023.

COUGHLAN, A. T. Marketing Channel Strategy. **Wiley International Encyclopedia of Marketing**, 2010. Chichester, UK: John Wiley & Sons, Ltd. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/9781444316568.wiem01038>>. .

DALVAND, B.; JAHANSHALOO, G.; LOTFI, H.; MALKHALIFE, M. R. **Using C4.5 Algorithm for Predicting Efficiency Score of DMUs in DEA**. 2014.

DING, L.; HAN, B.; WANG, S.; LI, X.; SONG, B. User-centered recommendation using US-ELM based on dynamic graph model in E-commerce. **International Journal of Machine Learning and Cybernetics**, v. 10, n. 4, p. 693–703, 2019. Springer Verlag.

DOMS, M. E.; JARMIN, R. S.; KLIMEK, S. D. Information technology investment and firm performance in US retail trade. **Economics of Innovation and New Technology**, v. 13, n. 7, p. 595–613, 2004.

DONTHU, N.; YOO, B. Retail productivity assessment using data envelopment analysis. **Journal of Retailing**, v. 74, n. 1, p. 89–105, 1998. JAI. Acesso em: 20/2/2023.

DOWNS, G. W.; ROCKE, D. M. Interpreting Heteroscedasticity. **American Journal of Political Science**, v. 23, n. 4, p. 816, 1979. Disponível em: <<https://www.jstor.org/stable/2110809?origin=crossref>>. .

DUBELAAR, C.; BHARGAVA, M.; FERRARIN, D. Measuring retail productivity: what really matters? **Journal of Business Research**, v. 55, n. 5, p. 417–426, 2002. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0148296300001600>>. Acesso em: 4/5/2023.

EHRET, M.; WIRTZ, J. Ownership of Co-creation Assets: Driving B2B Value Propositions in the Service Economy. **Journal of Creating Value**, v. 4, n. 1, p. 42–60, 2018. Disponível em: <<http://journals.sagepub.com/doi/10.1177/2394964318772146>>. .

FELICIANO-CESTERO, M. M.; AMEEN, N.; KOTABE, M.; PAUL, J.; SIGNORET, M. Is digital transformation threatened? A systematic literature review of the factors influencing firms' digital transformation and internationalization. **Journal of Business Research**, v. 157, 2023. Elsevier Inc.

FOMBY, T. B.; JOHNSON, S. R.; HILL, R. C. **Advanced Econometric Methods**. New York, NY: Springer New York, 1984.

G, R.; ASOKAN-AJITHA, A. Business customer experience in B2B2C service settings: a scale development. **Journal of Business and Industrial Marketing**, v. 38, n. 4, p. 886–904, 2023. Emerald Publishing.

GALINDO-MARTÍN, M. Á.; CASTAÑO-MARTÍNEZ, M. S.; MÉNDEZ-PICAZO, M. T. Digital transformation, digital dividends and entrepreneurship: A quantitative analysis. **Journal of Business Research**, v. 101, p. 522–527, 2019. Elsevier Inc.

HASKEL, J.; SADUN, R. Regulation and UK Retailing Productivity: Evidence from Microdata. **Economica**, v. 79, n. 315, p. 425–448, 2012.

HERBERT HEATON. **Productivity in Service Organizations**. New York: McGraw-Hill Inc, 1977.

HIGÓN, D. A.; BOZKURT, Ö.; CLEGG, J.; et al. The Determinants of Retail Productivity: A Critical Review of the Evidence. **International Journal of Management Reviews**, v. 12, n. 2, p. 201–217, 2010. John Wiley & Sons, Ltd. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1468-2370.2009.00258.x>>. Acesso em: 11/5/2023.

HUTT, M. D.; SPEH, T. W. **Business Marketing Management: B2B**. 2016.

KAMAKURA, W. A.; LENARTOWICZ, T.; RATCHFRORD, B. T. Productivity assessment of multiple retail outlets. **Journal of Retailing**, v. 72, n. 4, p. 333–356, 1996. JAI. Acesso em: 24/5/2023.

LEWIS, B.; BALLEK, M.; CRAIG, C.; HARRIS, V. **Driving productivity and growth in the UK economy**. London, 1998.

LUSCH, R. F.; SERPKENCI, R. R. Personal Differences, Job Tension, Job Outcomes, and Store Performance: A Study of Retail Store Managers. **Journal of Marketing**, v. 54, n. 1, p. 85–101, 1990. Disponível em: <<http://journals.sagepub.com/doi/10.1177/002224299005400106>>. .

MILLER, S.; STARTZ, R. Feasible Generalized Least Squares Using Machine Learning. **SSRN Electronic Journal**, 2018. Elsevier BV. Disponível em: <<https://papers.ssrn.com/abstract=2966194>>. Acesso em: 1/5/2022.

O'MAHONY, M.; DE BOER, W. Britain's Relative Productivity Performance: Has Anything Changed? **National Institute Economic Review**, v. 179, p. 38–43, 2002.

ROLLINS, J. B. **Foundational Methodology for Data Science**. 2015.

SHANKAR, V.; KUSHWAHA, T. Omnichannel marketing: Are cross-channel effects symmetric? **International Journal of Research in Marketing**, v. 38, n. 2, p. 290–310, 2021. Elsevier B.V.

SHIN, J.; KIM, Y. J.; JUNG, S.; KIM, C. Product and service innovation: Comparison between performance and efficiency. **Journal of Innovation & Knowledge**, v. 7, n. 3, p. 100191, 2022. Elsevier. Acesso em: 24/5/2023.

WHITE, H.; WHITE1, H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. , v. 48, n. 4, p. 817–838, 1980. Acesso em: 19/7/2022.

ZHANG, Z.; LI, J. Big-data-driven low-carbon management. **Big Data Mining for Climate Change**, p. 287–299, 2020. Elsevier. Acesso em: 26/2/2023.