data

Article

Quantifying Webpage Performance: A Comparative Analysis of TCP/IP and QUIC Communication Protocols for Improved Efficiency

Thyago Celso Cavalcante Nepomuceno [1,2,*], Késsia Thais Cavalcanti Nepomuceno [3], Fabiano Carlos da Silva [3] and Silas Garrido Teixeira de Carvalho Santos [3]

1 Department of Statistics, Federal University of Pernambuco, Av. Prof. Moraes Rego, 1235—Cidade Universitária, Recife 50670-902, Brazil
2 Department of Computer, Control, and Management Engineering Antonio Ruberti, Sapienza University of Rome, Piazzale Aldo Moro, 5, 00185 Roma, Italy
3 Center of Informatics, Federal University of Pernambuco, Av. Prof. Moraes Rego, 1235—Cidade Universitária, Recife 50670-902, Brazil; ktcn@cin.ufpe.br (K.T.C.N.); fcs3@cin.ufpe.br (F.C.d.S.); sgtcs@cin.ufpe.br (S.G.T.d.C.S.)
* Correspondence: thyago.nepomuceno@ufpe.br

Abstract: Browsing is a prevalent activity on the World Wide Web, and users usually demonstrate significant expectations for expeditious information retrieval and seamless transactions. This article presents a comprehensive performance evaluation of the most frequently accessed webpages in recent years using Data Envelopment Analysis (DEA) adapted to the context (inverse DEA), comparing their performance under two distinct communication protocols: TCP/IP and QUIC. To assess performance disparities, parametric and non-parametric hypothesis tests are employed to investigate the appropriateness of each website's communication protocols. We provide data on the inputs, outputs, and efficiency scores for 82 out of the world's top 100 most-accessed websites, describing how experiments and analyses were conducted. The evaluation yields quantitative metrics pertaining to the technical efficiency of the websites and efficient benchmarks for best practices. Nine websites are considered efficient from the point of view of at least one of the communication protocols. Considering TCP/IP, about 80.5% of all units (66 webpages) need to reduce more than 50% of their page load time to be competitive, while this number is 28.05% (23 webpages), considering QUIC communication protocol. In addition, results suggest that TCP/IP protocol has an unfavorable effect on the overall distribution of inefficiencies.

Keywords: data envelopment analysis; internet; communication protocols; efficiency analysis; websites; TCP/IP; QUIC; performance evaluation

check for updates

Citation: Nepomuceno, T.C.C.; Nepomuceno, K.T.C.; da Silva, F.C.; de Carvalho Santos, S.G.T. Quantifying Webpage Performance: A Comparative Analysis of TCP/IP and QUIC Communication Protocols for Improved Efficiency. *Data* **2023**, *8*, 134. https://doi.org/10.3390/data8080134

Academic Editor: Jamal Jokar Arsanjani

Received: 22 July 2023
Revised: 12 August 2023
Accepted: 15 August 2023
Published: 19 August 2023

1. Introduction

The Internet is essential for keeping people and companies working every day, and it takes most of our cycle time and existing social and cultural habits. It is estimated that over half of the world's population is using the Internet, increasing to over 69 percent among youth [1]. The perspective is that this number increases over time. The most common activity on the Internet is website navigation, which allows users to experience services, site features and retrieve different business information. Keeping the entire internet infrastructure with this expressive user number requires dynamic communication protocols and other fundamental technologies for internet operations.

Efficient webpages are essential because they directly impact the user experience, leading e-commerce companies to lose potential customers when they are unsatisfactory. Furthermore, the increasing number of users and data continuously leads the scientific community, companies, and researchers to investigate existing resources' performance on webpages, creating new technologies to meet the high demand and the increasing numbers of users. Communication protocols such as the TCP/IP and Quick UDP Internet Connections (QUIC) are essential for transmitting data and information following predefined rules. Although TCP/IP is still widely used worldwide, QUIC was adopted as an official internet standard protocol, which is expected to become more popular.

One of the main differences between TCP/IP and QUIC is that TCP/IP is a connection-oriented protocol that requires a three-way handshake to establish a connection between the client and server, while QUIC is designed to reduce the connection establishment latency by using a modified handshake mechanism that combines connection establishment and encryption setup into a single step. Another important feature of QUIC compared to TCP/IP is the capability for multiplexing, allowing multiple data streams to be transmitted over a single connection simultaneously. Depending on the built network, companies might guarantee a strategic advantage by meeting high peaks during specific seasons using different communication protocols.

The time required to load a page, i.e., Page Load Time (PLT), is one of the most critical metrics affecting the final user experience and satisfaction. The page load time influences the time a user will remain on the website. Users frustrated with low page load time tend to resent and not reaccess it, and websites with longer delays are considered less attractive and more difficult to find [2]. According to a recent survey by the Aberdeen Group [3], it takes only 3 s delay for about 21% of desktop web customers and about 20% of mobile web customers to leave websites, resulting in potential lost revenues. Page load time also influences the recommendation of users about the website.

Faster websites create satisfied users, and improving site speed also reduces operating costs, as recognized by Google, which uses the PLT in its algorithm for web searching ranking [4], directing the users to websites with a lower load time. The importance of page load time on the Internet tends to increase more due to the massive amount of data being processed as page size and objects over time. This work evaluates the efficiency of the world's most-accessed webpages in recent years based on an inverse perspective of Data Envelopment Analysis (DEA) frontier estimation. The DEA is a linear programming technique that does not require the predefining of a functional form (i.e., they are non-parametric by nature) designed by Charnes, Cooper, and Rhodes [5] in 1978 for multiple output/input configurations.

The DEA methodology for measuring the technical efficiency of decision units has increased exponentially in the number of methodological contributions, empirical applications, and computational development over the years [6–8]. Data Envelopment Analysis estimates a non-parametric efficiency frontier based on the determination of production possibilities in which Decision-Maker Units (DMUSs) are compared to each other in order to evaluate each decision unit efficiency as a ratio between the used resources and obtained results, which may contain the influence of exogenous factors [8,9]. The inverse perspective for the DEA frontier estimation considers outputs to be reduced (as bad outputs) and inputs to be increased proportionally.

Evaluating network communications and webpage ranking is an important subject [10–13]. Many DEA applications and evaluations of efficiency have settled on energy sectors [14–16], services [17,18], communications [19–21], and retail logistics and e-commerce domains [22]. The evaluation of websites, pages, and web services has been a point of interest in the literature with applications that support the construction of evaluation systems based on Alexa Rankings and users experience, reduction of page load time, and impact in terms of objects, images, and servers [23,24]. Meanwhile, to the best of our knowledge, despite the evident relevance, websites lack non-parametric efficiency assessments. Two instances of the use of DEA on webpages can be consulted in evaluating Jordan University webpages based on the perspectives of usability, design, and performance [25] and the operational efficiency of pages for e-commerce with a focus on service quality [26].

This work aims to measure the technical efficiency of websites by the ratio of website resources to page load time. This methodology is explained in Section 2.2. According to technical efficiency measures (the ratio of used resources to the produced results), the evaluation ranks webpages under two important communication protocols: TCP/IP and QUIC. Both are the options currently used by available web browsers. The page size and number of objects are considered as production inputs (website resources) used to pro- duce the output Page Load Time (PLT). The bigger the page size or the number of objects (inputs), the longer the webpage's loading (output). Because we assume there are no additional costs by including more objects or increasing the page size, and the page load time is considered a bad output, DEA is applied in an inverse frontier perspective: efficient pages are those with the lowest PLT proportional to the number of objects and size. The inefficient pages aim for efficient ones for performance benchmarks. The results report the best-performing pages and quantitative potential for improvements to reach the efficient frontier.

The paper is organized as follows: the next section describes the theoretical methodology of the Data Envelopment Analysis adjusted to the context and the methodology used in conducting the experiments, that is, the evaluation scenario, the tools used, and how the experiments were performed. Section 3 presents and discusses the website's efficiency results and differences from both communication protocol's perspectives. The last section concludes by summarizing remarks and suggestions for future works.

2. Materials and Methods

The collection and analysis of experimental data were conducted based on a methodology composed of four main activities: (a) measurement planning, defining how and which metrics will be used; (b) measurements in a controlled environment, evaluating the configuration and operation of webpages; (c) Data Envelopment Analysis application and statistical treatments; and (d) discussion of the results. The adapted DEA methodology, the configuration of the experiments, metrics, and factors are detailed in the following subsections.

Common and consolidated tools by the software community were used in the experiments. For the execution of the experiments, the environment was configured to correspond to today's internet as much as possible. As discussed by [13], using actual internet links can influence the connection dynamics, affecting the evaluation and producing unfair comparisons. Not rarely do the characteristics of the specific link differ from the typical conditions for some websites. In addition, using Google servers may be beneficial for QUIC-based processing (see some instances of such protocol performance evaluation in [27–36]). Based on the literature research, we established several configurations influencing and representing the real-world internet. The performed settings are described in Section 2.1. Section 2.2 discusses the adapted Data Envelopment Analysis model applied to this website assessment.

2.1. Testbed, Experiments and Factors

The following described environment here was designed to reflect the actual conditions of Internet traffic in the most reliable way possible. We considered the 100 most accessed websites in the world according to Alexa website ranking [37] to perform the experiments. The top 100 sites in Alexa provide a good internet representation because it displays the most accessed websites globally. We used the Google Big Query tool to retrieve this analysis's inputs and output. However, it was not possible to collect information on the number of objects and page size for 18 pages from the 100 most accessed websites due to, among other factors, URL redirections or because they were no longer in operation. In addition, for these and other factors, some pages did not return the loading time during the access.

Eighty-two of the original units have all the information required for the analysis. This is sufficient to avoid the curse of dimensionality [38] and maintain the robustness of the analysis containing two inputs and one product. For this reason, we opted to work with 82 pages out of the 100 most accessed pages from which it was possible to extract all information for the analysis. The removed pages, with their respective positions based on Alexa's report, were: yahoo.com (5th); t.co (34th); paypal.com (39th); google.com.tr (61st); office.com (63rd); diply.com (66th); bongacams.com (80th); googleusercontent.com (82nd); jd.com (84th); chinadaily.com.cn (89th); 360.com (90th); coccoc.com (91st); adnetworkperformance.com (93rd); haosou.com (95th); adobe.com (96th); china.com (97th); microsoftonline.com (99th); and whatsapp.com (100th).

The environment was set up on virtual machines using VMWare Player to reproduce the pages, which allowed paralleling several executions and ensuring the same configuration for all identical virtual machines. The operating system for these machines was Ubuntu 16.04. The browser was Google Chrome 54.0.28.40.100. The websites belong to different servers. Google BigQuery tool was used to understand how the content is organized and find the number of objects, which contains an extensive database of webpage information [39]. The Alexa pages were uploaded on Google BigQuery, and, through consultations, the inputs used in the data envelopment analysis experiments were retrieved, i.e., the number of objects and the size of these objects (page size) measured in bytes.

Creating a script to record and reproduce the pages using Mahimahi was necessary for the output. This tool records websites and reproduces them under an emulated network condition [40]. Using Mahimahi through virtual machines, it was possible to reproduce the websites and capture the page load time output. The webpages were loaded with 100 milliseconds RTT and 1% packet loss for both protocols. TCP/IP with HTTP 1.1 application protocol and QUIC with HTTP 2. These RTT and packet loss values were chosen based on the average values in [41]. The HAR file was captured from the page and the HAR file; the onLoad event was extracted to be used as page load time in the experiments.

Therefore, to carry out the data envelopment analysis, 82 sites were evaluated, considering the number of objects and page size as inputs and the OnLoad metric for page load time as the output. The mentioned inputs were chosen due to the capacity to produce (increasing) page load time, directly affecting the performance of websites performing well or poorly compared to the others. This analysis makes it possible to infer the best-performing pages needing improvements to reach the efficiency frontier.

2.2. Data Envelopment Analysis

The traditional concept of technical efficiency, according to Shephard [42], Fare and Lovell [43], and Charnes and associates [5], is related to the maximum radial contraction or expansion that a decision-making unit (industries, companies, groups, people, machines, technologies) can obtain for production resources or products, respectively. Following this reasoning, a technically inefficient unit can become efficient by increasing its results (output) or reducing its resources (inputs), keeping similar production levels. In the first case, we have output-oriented models; in the second, input-oriented models. The efficiency measurement using the DEA technique comes from the ratio of outputs produced to the inputs consumed in a production process. It is a linear optimization mathematical programming technique with thousand of empirical applications and theoretical contributions over the past decades in all sectors of economic activity [6–8]. In this methodology, we can identify efficiency scores and potentials for improvement by solving the following dual form (considering the input-oriented case):

$$
\begin{gathered}
min\ \theta \\
Subject\ to: \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{io} | i = 1, 2, 3, \ldots, m \\
\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro} | r = 1, 2, 3, \ldots, s \\
\sum_{j=1}^{n} \lambda_j = 1 | j = 1, 2, 3, \ldots, n
\end{gathered}
\tag{1}
$$

$$\lambda_j \geq 0$$

where x_{ij} is the input vector i = 1, 2, . . ., m used by each decision unit j; y_{rj} is the output vector r = 1, 2, . . ., s produced by each decision unit j; λ_j is the dual multiplier which translates the weighting contribution by each unit in the linear optimization and construction of the production frontier; and θ is the technical efficiency, which multiplied by the input vector of unit "o" under analysis (x_{io}) offers the optimal reduction which the unit must pursue to attain efficiency.

The linear formulation results in an efficiency score e = $(1 - \theta)$, which varies from 0 to 1. Pages scoring 1 are technically efficient and benchmarks for best practices in reducing internet latency. Inefficient pages, for instance, scoring 0.8, need to improve their score by 0.2 to attain efficiency. This can be possible by expanding 20% of their results (output), keeping the same level of resources (input), or by reducing 20% of their resources, keeping the same results. Through a weight optimization process in the linear programming model (1), the methodology allows the construction of a non-parametric surface (frontier) by pairwise comparisons involving the data [32].

Figure 1a illustrates this reasoning for a general model under constant returns to scale. There are six efficient units (located along the straight line representing the efficiency frontier) and eleven inefficient units (enveloped by the linear combination of those six). The inefficient unit "3" can become efficient by either reducing its input proportionally to the same level as the efficient unit "1", keeping the same output, or by expanding its output to the same level as the efficient unit "2", keeping the same input usage (radial contractions and expansions).

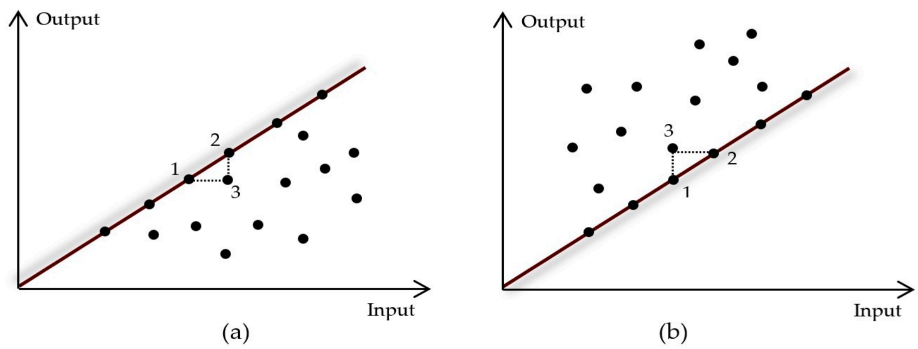

Figure 1. Schematic representations of efficiency frontiers under constant returns to scale. Panel (**a**): standard CRS frontier enveloping inefficient units using one input to produce one output. Panel (**b**): inverse CRS frontier enveloped by inefficient units using one free-disposal and cost-free input to produce one bad output.

Because the context of this evaluation regards a bad output, we aim at optimizing the result by reducing to the maximum possible page load time, PLT. Furthermore, because the inputs are assumed to be cost-free, we aim to obtain the maximum possible use of graphic and digital resources (size and number of objects), given the loading time restrictions. Thus, the linear programming input-oriented formulation defined in (1) can be adjusted to an inverse perspective:

$$max\ \varphi$$

$$Subject\ to: \sum_{j=1}^{n} \lambda_j y_{rj} \leq y_{ro} | r = 1,\ 2,\ 3,\ \ldots,\ s$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} \geq \varphi x_{io} | i = 1,\ 2,\ 3,\ \ldots,\ m \tag{2}$$

$$\sum_{j=1}^{n} \lambda_j = 1\ | j = 1,\ 2,\ 3,\ \ldots,\ n$$

$$\lambda_j \geq 0$$

where x_{ij} is the input vector i = 1, 2, . . ., m used by each decision unit j; y_{rj} is the output vector r = 1, 2, . . ., s produced by each decision unit j; λ_j is the dual multiplier which translates the weighting contribution by each unit in the linear optimization and construction of the production frontier; and φ is the technical inversed efficiency varying from 0 to 1, which multiplied by the input vector of unit "o" under analysis (x_{io}) offers the optimal expansion which the unit must pursue to attain efficiency. The interpretation of the parameters to the data in context is as follows: the efficient webpages in this model report φ = 1, which means that they operate with a low page load time using more objects and other graphical and digital resources compared to the others. In other words, they maximize the use of resources φx_{io} (see the functional and second restriction in (2)), keeping the same level of page load time y_{ro} (see the first restriction in (2)).

Alternatively, considering a slack-based additive model with both input and output orientation [44,45]:

$$min\ \varphi - e\left(\sum_{i=1}^{m} s_i + \sum_{r=1}^{s} s_r\right)$$

$$Subject\ to: \sum\nolimits_{j=1}^{n} \lambda_j y_{rj} + s_r = y_{ro} | r = 1,\ 2,\ 3,\ \ldots,\ \mathrm{s}$$

$$\sum\nolimits_{j=1}^{n} \lambda_j x_{ij} - s_i = \varphi x_{io} | \mathrm{i} = 1,\ 2,\ 3,\ \ldots,\ \mathrm{m} \tag{3}$$

$$\lambda \mathrm{j} \geq 0,$$

$$\sum\nolimits_{j=1}^{n} \lambda_j = 1 \quad | \mathrm{j} = 1,\ 2,\ 3,\ \ldots,\ n$$

$$\lambda \mathrm{j} \geq 0,$$

where s_i and s_r are the production slacks to be contracted or expanded.

This model provides the construction of an inverted non-parametric efficiency frontier enveloped by the inefficient unit's data, as represented by Figure 1b. The efficiency score for each DMU ranging from 0 to 1 reflects the need for efficiency gains. Slacks in this model are interpreted in the opposite direction: Page Load Time (PLT) slacks measure what the page under evaluation wants to reduce instead of increasing (output). Slacks for page size or the number of objects represent what the page under evaluation is expected to expand compared to benchmark with similar technology. In the illustration of Figure 1b, the inefficient unit "3" may attain efficiency either by reducing the bad output to the efficient level of "1" or by expanding the page cost-free resources to the efficient level of unit "2".

3. Results and Discussion

Table 1 reports the inputs, output, and respective efficiency scores (in percentage) derived from efficiency analysis of 82 out of the world's top 100 most accessed websites. The PLT output and efficiency scores regard the TCP/IP (fourth and sixth) and QUIC (fifth and seventh). The first column presents the pages (Decision-Making Units—DMUs) ranked by the Efficiency column (built from the maximum value between the sixth and seventh columns). The number of objects and page size (in bytes) are reported as inputs on the second and third columns.

Table 1. Inputs, outputs, and efficiency scores.

Websites (DMUs)	Objects	Page Size	PLT (TCP/IP)	PLT (QUIC)	Efficiency (TCP/IP)	Efficiency (QUIC)	Higher Efficiency
sina.com.cn	**496**	**19,069,095**	$\mathbf{1.66 \times 10^4}$	$\mathbf{1.27 \times 10^4}$	**100.00%**	**100.00%**	**100.00%**
rakuten.co.jp	**808**	**10,521,295**	$\mathbf{1.08 \times 10^4}$	$\mathbf{1.33 \times 10^4}$	**100.00%**	**100.00%**	**100.00%**
cntv.cn	**130**	**1,021,743**	$\mathbf{2.63 \times 10^3}$	$\mathbf{3.18 \times 10^3}$	**100.00%**	**100.00%**	**100.00%**
bing.com	**56**	**941,086**	$\mathbf{1.72 \times 10^3}$	$\mathbf{5.35 \times 10^3}$	**100.00%**	**92.11%**	**100.00%**
amazon.com	**199**	**4,639,244**	$\mathbf{8.59 \times 10^3}$	$\mathbf{1.31 \times 10^4}$	**100.00%**	**89.98%**	**100.00%**
xinhuanet.com	**92**	**14,845,093**	$\mathbf{7.16 \times 10^3}$	$\mathbf{5.57 \times 10^3}$	**88.77%**	**100.00%**	**100.00%**
cnn.com	**17**	**678,982**	$\mathbf{1.59 \times 10^4}$	$\mathbf{6.66 \times 10^3}$	**59.82%**	**100.00%**	**100.00%**
imdb.com/	**17**	**670,175**	$\mathbf{1.51 \times 10^4}$	$\mathbf{7.30 \times 10^3}$	**57.58%**	**100.00%**	**100.00%**
xvideos.com	**20**	**263,552**	$\mathbf{2.66 \times 10^4}$	$\mathbf{4.07 \times 10^3}$	**32.27%**	**100.00%**	**100.00%**
baidu.com	91	1,507,173	3.24×10^3	3.34×10^3	61.28%	95.37%	95.37%
naver.com	101	1,537,049	6.73×10^3	5.98×10^3	76.52%	93.07%	93.07%
msn.com	92	758,056	2.87×10^3	9.67×10^3	91.51%	65.83%	91.51%
mail.ru	15	641,062	9.53×10^3	7.01×10^3	55.14%	90.74%	90.74%
amazon.in	136	2,760,955	1.16×10^4	2.66×10^4	90.58%	50.21%	90.58%
pornhub.com	128	2,577,879	7.62×10^3	1.23×10^4	90.18%	76.70%	90.18%
flipkart.com	254	5,669,366	9.50×10^3	9.74×10^3	73.37%	89.79%	89.79%
hao123.com	385	1,573,258	8.89×10^3	8.21×10^3	68.20%	88.31%	88.31%
sohu.com	174	2,624,886	1.23×10^4	1.45×10^4	85.59%	87.61%	87.61%
tumblr.com	112	1,621,744	2.51×10^4	1.46×10^4	74.19%	87.48%	87.48%
amazon.co.jp	132	789,647	1.75×10^4	1.46×10^4	61.49%	87.00%	87.00%
ebay.com	302	6,225,062	9.33×10^3	1.01×10^4	77.68%	86.71%	86.71%
amazon.de	18	576,698	1.29×10^4	1.90×10^4	83.70%	66.97%	83.70%
google.co.uk	96	869,881	5.63×10^3	3.86×10^3	61.06%	82.37%	82.37%
google.de	258	5,060,789	5.49×10^3	4.01×10^3	62.63%	79.36%	79.36%
youtube.com	131	1,589,924	6.62×10^3	2.11×10^4	77.98%	47.29%	77.98%
gmw.cn	36	487,700	1.44×10^4	1.11×10^4	53.95%	76.63%	76.63%
amazon.co.uk	182	2,150,533	1.41×10^4	2.39×10^4	76.35%	53.41%	76.35%
nicovideo.jp	84	2,040,864	7.75×10^3	9.01×10^3	66.67%	76.19%	76.19%
pixnet.net	47	1,239,330	1.80×10^4	1.77×10^4	59.84%	71.97%	71.97%
kat.cr	307	4,408,514	1.09×10^4	1.11×10^4	63.85%	70.77%	70.77%
github.com	165	1,504,194	8.27×10^3	8.01×10^3	62.29%	69.51%	69.51%
twitter.com	68	1,423,637	6.08×10^3	4.65×10^3	43.25%	68.40%	68.40%
360.cn	148	1,869,267	9.75×10^3	1.33×10^4	68.18%	66.04%	68.18%
microsoft.com	102	1,257,262	1.03×10^4	2.00×10^4	66.95%	52.55%	66.95%
dropbox.com	20	53,272	2.16×10^4	1.91×10^4	49.75%	66.86%	66.86%
xhamster.com	15	641,922	7.84×10^3	9.58×10^3	55.40%	66.44%	66.44%
qq.com	15	647,486	1.84×10^4	1.81×10^4	56.48%	65.80%	65.80%
tianya.cn	54	1,279,964	8.63×10^3	4.87×10^3	30.77%	65.39%	65.39%
soso.com	23	276,948	3.71×10^3	4.87×10^3	46.30%	65.33%	65.33%
imgur.com	14	1,638,025	6.66×10^3	1.38×10^4	65.28%	55.44%	65.28%
go.com	110	940,945	1.39×10^4	9.65×10^3	45.88%	64.06%	64.06%
outbrain.com	15	638,836	9.36×10^3	1.12×10^4	55.08%	61.91%	61.91%
taobao.com	18	679,504	1.52×10^4	1.43×10^4	56.66%	61.13%	61.13%
google.fr	182	3,993,262	5.70×10^3	1.07×10^4	60.22%	52.11%	60.22%
google.com.mx	17	672,061	5.87×10^3	8.80×10^3	58.55%	50.59%	58.55%
google.pl	17	669,256	6.03×10^3	8.42×10^3	56.96%	52.51%	56.96%
pinterest.com	141	1,669,451	1.14×10^4	9.79×10^3	48.48%	56.87%	56.87%
fc2.com	23	459,604	1.15×10^4	7.01×10^3	37.77%	56.68%	56.68%
alibaba.com	147	2,861,726	9.13×10^3	1.41×10^4	56.45%	47.29%	56.45%
yandex.ru	184	387,564	7.46×10^3	1.42×10^4	56.25%	44.77%	56.25%
google.it	16	649,383	6.12×10^3	9.60×10^3	56.17%	53.09%	56.17%
tmall.com	36	571,077	1.66×10^4	9.92×10^3	29.10%	55.89%	55.89%
google.co.in	15	647,513	6.15×10^3	9.44×10^3	55.82%	50.15%	55.82%
google.co.id	19	642,695	6.20×10^3	8.82×10^3	55.43%	49.91%	55.43%

Table 1. *Cont.*

Websites (DMUs)	Objects	Page Size	PLT (TCP/IP)	PLT (QUIC)	Efficiency (TCP/IP)	Efficiency (QUIC)	Higher Efficiency
wikipedia.org	206	2,144,843	3.18×10^3	7.82×10^3	54.05%	40.70%	54.05%
apple.com	35	129,214	2.35×10^4	1.82×10^4	43.24%	53.65%	53.65%
google.es	57	512,902	6.97×10^3	1.01×10^4	49.28%	53.09%	53.09%
google.ru	208	4,239,393	6.51×10^3	1.02×10^4	52.74%	50.21%	52.74%
google.com.br	63	585,770	6.59×10^3	1.06×10^4	52.10%	50.59%	52.10%
google.co.jp	220	2,314,736	6.66×10^3	8.93×10^3	51.56%	50.73%	51.56%
google.com.hk	14	525,033	6.57×10^3	8.63×10^3	51.39%	47.18%	51.39%
google.ca	17	679,515	6.71×10^3	8.42×10^3	51.22%	50.08%	51.22%
aliexpress.com	126	2,187,887	1.13×10^4	2.20×10^4	50.43%	39.77%	50.43%
google.com.au	15	645,025	6.88×10^3	8.58×10^3	49.93%	50.39%	50.39%
yahoo.co.jp	20	141,242	8.78×10^3	1.04×10^4	49.49%	50.20%	50.20%
google.com	39	448,042	8.69×10^3	1.10×10^4	39.52%	50.04%	50.04%
facebook.com	41	578,970	5.81×10^3	6.39×10^3	47.73%	49.78%	49.78%
wordpress.com	59	579,111	8.00×10^3	6.48×10^3	32.87%	49.12%	49.12%
ok.ru	53	2,069,649	9.32×10^3	1.17×10^4	36.85%	47.79%	47.79%
reddit.com	53	474,662	8.81×10^3	9.43×10^3	46.46%	46.46%	46.46%
live.com	10	162,069	5.07×10^3	6.88×10^3	33.90%	46.22%	46.22%
weibo.com	16	640,444	1.22×10^4	1.44×10^4	46.05%	46.05%	46.05%
craigslist.org	26	288,626	4.46×10^3	6.97×10^3	38.56%	45.66%	45.66%
instagram.com	16	132,472	1.00×10^4	7.78×10^3	34.29%	44.98%	44.98%
vk.com	40	306,612	8.97×10^3	7.17×10^3	38.30%	44.39%	44.39%
stackoverflow.com	31	1,368,820	6.12×10^3	1.10×10^4	42.96%	28.94%	42.96%
ask.com	82	1,332,187	5.08×10^3	9.95×10^3	41.07%	31.99%	41.07%
blogger.com	14	78,710	7.90×10^3	8.76×10^3	25.00%	36.32%	36.32%
blogspot.com	14	78,667	8.40×10^3	9.31×10^3	25.00%	34.19%	34.19%
netflix.com	18	325,990	1.51×10^4	1.70×10^4	34.18%	32.78%	34.18%
onclickads.net	12	152,909	5.26×10^3	1.00×10^4	32.68%	31.74%	32.68%
linkedin.com	13	702,609	8.40×10^3	1.03×10^4	31.91%	31.91%	31.91%

Nine websites are considered efficient from the point of view of at least one of the protocols: sina.com.cn, rakuten.co.jp, cntv.cn, bing.com, amazon.com, xinhuanet.com, cnn.com, imdb.com, and xvideos.com. The remaining 73 websites are considered inefficient under this methodology. About 80.5% of all units (66 webpages) need to reduce more than 50% of their page load time proportionally to the expansion of resources conditional to TCP/IP, compared to about 28.05% (23 webpages) conditional to QUIC. This may suggest that different communication protocols may affect the empirical distribution of inefficiencies. Nepomuceno et al. [10,46] argued that when exogenous factors have an unfavourable effect on the overall efficiency, the probability for a given unit to be located far from the efficient frontier (being less efficient) increases.

Conversely, when exogenous factors have a favorable effect on the overall efficiency, the probability for a given unit to be located far from the efficient frontier (being less efficient) decreases. In addition to the empirical distribution of the inefficiencies, about half the efficient units do not remain efficient for both communication protocols perspectives. Considering the different communication protocols as exogenous factors for the production of load time, this may suggest that different communication protocols may also affect the shape of the attainable frontier. Testing the separability condition as proposed by Daraio, Simar, and Wilson [47] may aid the broad understanding of communication protocols' influence on the performance measurement of websites.

The overall additive technical efficiency is reinforced in three website units. For sina.com.cn, rakuten.co.jp, cntv.cn, we observe that the webpages are technically efficient for both protocols' applications. This confirms the efficiency of these webpages. For the additional efficient and inefficient webpage units, there is a higher efficient rate favoring the

QUIC communication protocol compared to the TCP/IP communication protocol, which supports the remark that most webpages are suitable for the most recent protocol.

Figure 2 illustrates the dispersion plot for investigating potential relations between TCP/IP and QUIC efficiency scores. A regression red line is included to facilitate the visualization. There is some weak evidence for a positive relationship between the efficiency scores conditional to the communication protocols (correlation = 0.6323), which means they walk in the same direction most of the time. Particular cases reporting high differences for specific page units may present an interesting topic for additional investigations.

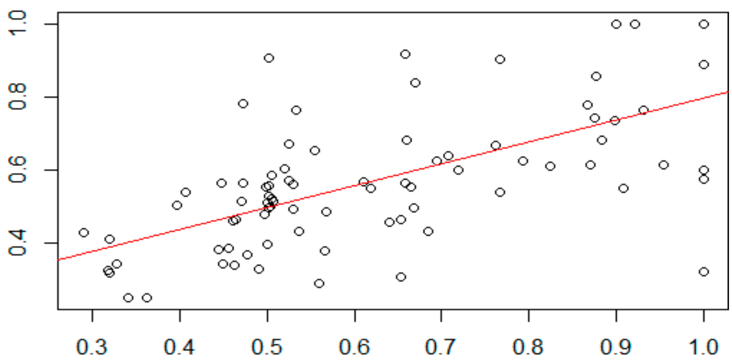

Figure 2. TCP/IP vs. QUIC dispersion graph. Pearson's correlation = 0.6322823.

Different communication protocols may have different effects on the empirical distribution of inefficiencies. This assertion has the additional statistical support of parametric and non-parametric hypothesis tests performed over the efficiency score comparing both approaches. Assuming equal variances and paired data, both two-sample Mann-Whitney and Student's *t*-tests reject the null hypothesis of median and mean equality favouring the alternative for a negative difference between TCP/IP efficiency and QUIC efficiency, which means that the TCP/IP protocol has an unfavourable effect over the overall distribution of inefficiencies. In other words, keeping all the other factors constant, the probability for the same webpage to be located far from the efficient frontier (i.e., less efficient) increases using the TCP/IP protocol compared to QUIC. Figure 3 illustrates this comparison using notched boxplots to visualize potential differences.

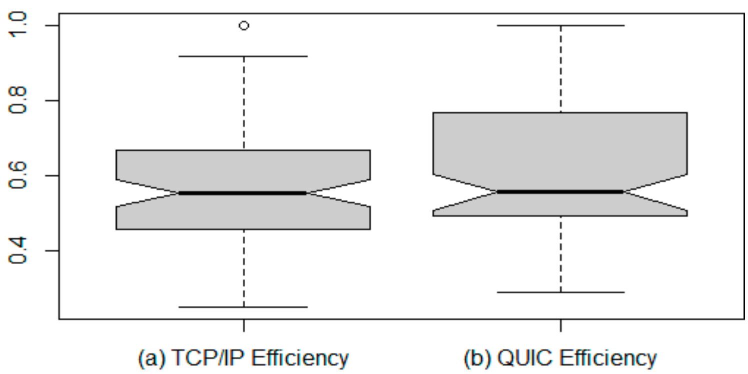

Figure 3. Boxplots and hypothesis tests.

Percentage changes in the combined input and output values measured by the inefficiency score provide the optimal contraction and expansion for each page in order to attain efficiency. This potential for improvement on the inputs and outputs is reported in Table 2 for the 20 least QUIC-efficient webpages considering both communication protocols' perspectives. For instance, a webpage such as ask.com, which is 41.07% TCP/IP efficient and 31.98% QUIC-efficient, would require to expand about 59% objects and size (13.55 and 163,201.56 bytes) and reduce about 59% page load time (2.99×10^3) considering the

TCP/IP perspective. The same page in the QUIC perspective would require to expand of about 68% objects and size (15.64 and 188,359.93 bytes) and reduce about 68% page load time (6.76×10^3). This performance improvement can be obtained through benchmarking best practices from the efficient pages with similar inputs to output configuration.

Table 2. Potential for improvements (for selected units).

DMU	Potential for Improvements (TCP/IP Efficiency)			Potential for Improvements (QUIC Efficiency)		
	Objects	Page Size	PLT (TCP/IP)	Objects	Page Size	PLT (QUIC)
stackoverflow.com	14.83	164,619.59	3.49×10^3	18.48	205,097.19	7.81×10^3
onclickads.net	8.08	102,937.00	3.54×10^3	8.19	104,376.81	6.84×10^3
linkedin.com	12.26	221,982.01	5.72×10^3	12.26	221,982.01	6.99×10^3
ask.com	13.55	163,201.56	2.99×10^3	15.64	188,359.93	6.76×10^3
netflix.com	8.56	462,425.94	9.92×10^3	8.74	472,271.67	1.14×10^4
blogspot.com	10.50	59,000.26	6.30×10^3	9.21	51,774.98	6.13×10^3
blogger.com	10.50	59,032.52	5.92×10^3	8.92	50,123.60	5.58×10^3
aliexpress.com	33.71	705,763.49	5.60×10^3	40.96	857,464.81	1.33×10^4
wikipedia.org	9.19	24,477.28	1.46×10^3	11.86	31,592.93	4.64×10^3
vk.com	24.06	276,454.05	5.53×10^3	21.69	249,153.84	3.99×10^3
yandex.ru	27.56	256,274.43	3.27×10^3	34.80	323,546.85	7.85×10^3
instagram.com	15.11	302,003.56	6.58×10^3	12.65	252,863.42	4.28×10^3
craigslist.org	6.14	99,578.41	2.74×10^3	5.43	88,069.98	3.79×10^3
weibo.com	59.35	507,676.84	6.56×10^3	59.35	507,676.84	7.78×10^3
live.com	10.58	87,568.65	3.35×10^3	8.60	71,240.70	3.70×10^3
reddit.com	28.38	254,152.60	4.72×10^3	28.38	254,152.64	5.05×10^3
google.com.hk	6.81	255,239.57	3.19×10^3	7.40	277,342.79	4.56×10^3
alibaba.com	20.47	539,765.86	3.98×10^3	24.77	653,243.03	7.42×10^3
youtube.com	28.85	350,162.01	1.46×10^3	69.05	838,010.56	1.11×10^4
ok.ru	25.89	365,618.48	5.89×10^3	21.41	302,284.75	6.08×10^3

Figure 4 illustrates the non-parametric inverse efficient production frontier considering the TCP/IP (Figure 4a) and QUIC (Figure 4b) communication protocols constructed through the additive model under both orientations. The output (Y) axis represents the page load time in milliseconds. The input (X) axis represents the combination of the two inputs, number of objects, and page size. On both panels, frontiers are non-convex, and a VRS frontier is also included for comparison purposes concerning the piecewise additive efficiency frontier representation. The differences in the empirical distribution of inefficiencies and the shape of the production frontier suggest testing the separability condition of exogenous factors and using semi-parametric conditional frontier estimators [9,46–50].

In the visual representations, for one of the 77 TCP/IP inefficient pages or for one of the 75 QUIC inefficient pages (points on the left of the piecewise frontiers) to reach the efficiency frontier, it is either necessary to move to the right increasing the resources (objects and size) or move down decreasing the results (page load time) or a combination of both. Doing that, those pages go towards one of the five TCP/IP efficient benchmark points (amazon.com, sina.com.cn, bing.com, rakuten.co.jp, cntv.cn) located at the frontier of Figure 4a or towards one of the seven QUIC efficient points (sina.com.cn, akuten.co.jp, cntv.cn, xinhuanet.com, cnn.com, imdb.com, xvideos.com) on the boundary of the efficiency frontier of Figure 4b. This illustration of the inverse efficient production frontiers of TCP/IP and QUIC communication protocols using an additive model highlights the implications for efficiency enhancement and the necessity to consider various adjustments for inefficient pages, such as adaptability, multiplexing, and retransmission of lost packets.

The overall potential for improvements from the results (by summing all potential expansion in the resources and reduction in the page load time for all inefficient pages) reports the possibility of additional 1845.035 page objects and 29,540,758.37 bytes (29.54 MB) can be added to inefficient webpages' capabilities, and about 2.48521×10^5 (4.14 min) TCP/IP page load time and 2.97142×10^5 (4.95 min) QUIC page load time can be saved in this request-response cycle. From a practical managerial perspective, such performance improvement can be obtained through benchmarking networking strategies from efficient pages with similar production configurations.

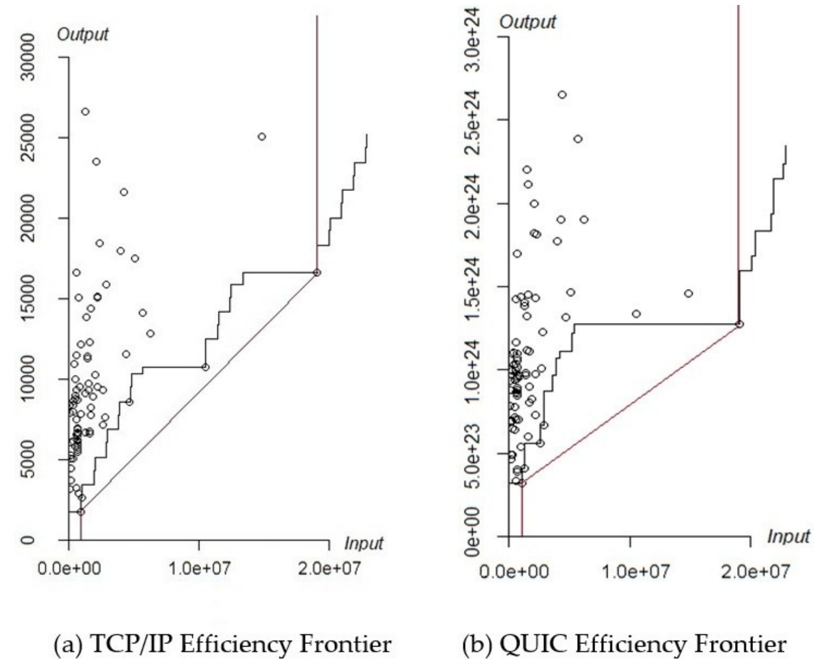

Figure 4. Input-output additive hyperbolic graph-orientated efficiency frontiers. Panel (**a**): TCP/IP inverse hyperbolic graph-orientated efficiency frontier and an equivalent VRS efficiency frontier (dark red). Panel (**b**): QUIC inverse hyperbolic graph-orientated efficiency frontier and an equivalent VRS efficiency frontier (dark red).

4. Conclusions

This work adapted the traditional non-parametric measure for technical efficiency to the context of website evaluation where resources are cost-free and the product, the page load time, is a bad output to be reduced instead of increased. This inverse frontier methodology assessed the efficiency of some of the world's most-accessed websites from two communication protocol perspectives: TCP/IP and QUIC. The efficiency measures and hypothesis tests support the assertion that different communication protocols affect the empirical distribution of inefficiencies and the shape of the production frontier. For example, in the case of QUIC, the probability for a specific website to be located far from the efficient frontier (being less efficient) decreases, keeping everything else constant.

This study has adapted the conventional non-parametric efficiency measure to the unique context of websites' evaluation, where resources are considered cost-free, and the objective is to minimize the detrimental output of page load time. Through the novel inverse frontier methodology, the technical efficiency of globally accessed websites has been assessed within TCP/IP and QUIC communication protocols. The obtained efficiency metrics and hypothesis tests underscore the significant impact of different communication protocols on inefficiency distribution and production frontier shape. Notably, in the case of QUIC, the likelihood of websites deviating considerably from the efficient frontier diminishes while holding other factors constant.

Furthermore, the study provides crucial efficiency metrics, benchmarks, avenues for enhancement, and rankings for comparative analysis. The analysis reveals substantial potential for improvement, suggesting the possibility of augmenting the capabilities of inefficient webpages by adding a significant number of objects and reducing page load time. These improvements translate to more efficient TCP/IP and QUIC page load times. The insights garnered from this work contribute significantly to enhancing website efficiency and user experience through informed network strategy benchmarking and replication.

Including specific setups for connection, jitter can be an interesting extension of the current analysis to investigate effects such as out-of-order packet delivery, increased latency, and even dropped packets, which can negatively impact real-time applications. Some interesting network strategies adopted by efficient websites can be benchmarked and reproduced by inefficient sites without reducing page size, objects, elements compression, or cache-reused content peers. Some instances are using Content Delivery Networks to reduce the distance between servers and users: minification (removing unnecessary code, white spaces, and comments); applying the minimum number of domains and non-nested tables; and reducing markups, redirects, and inline scripts.

Another interesting strategy, which does not affect the efficiency of the website but affects the user satisfaction, is optimizing the loading order of the webpage components, such as downloading first text content and any JavaScript required to display visual information so that users may have an apparent quicker response during the page loading. We hope to address many of those discussions in future evaluations.

Author Contributions: Conceptualization, T.C.C.N. and K.T.C.N.; methodology, T.C.C.N. and K.T.C.N.; software, T.C.C.N. and K.T.C.N.; validation, T.C.C.N., K.T.C.N. and S.G.T.d.C.S.; formal analysis T.C.C.N. and K.T.C.N.; investigation, T.C.C.N. and K.T.C.N.; resources, T.C.C.N. and K.T.C.N.; data curation, T.C.C.N. and K.T.C.N.; writing—original draft preparation, all authors; writing—review and editing, all authors; visualization, T.C.C.N. and K.T.C.N.; supervision, T.C.C.N.; project administration, T.C.C.N.; funding acquisition, T.C.C.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Universidade Federal de Pernambuco (UFPE), granted by Edital de Apoio ao Pesquisador vinculado aos Programas de Pós-Graduação da UFPE number Edital PROPG nº 06/2022. The APC was funded by Universidade Federal de Pernambuco Edital PROPG nº 06/2022.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are reported in this manuscript.

Acknowledgments: Authors acknowledge the support from family members, colleagues from the Networking and Telecommunications Research Group (GPRT-UFPE), and the Brazilian Coordination for the Improvement of Higher Education (CAPES).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. International Telecommunication Union. Measuring Digital Development: Facts and Figures. Available online: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf (accessed on 22 July 2023).
2. Selvidge, P. How Long Is Too Long to Wait for a Website to Load. *Usability News* **1999**, *1*, 1–3.
3. Aberdeen Group. Gone in Three Seconds: Your Customers Have no Patience for a Slow Website. Available online: https://www.aberdeen.com/techpro-essentials/gone-in-three-seconds-your-customers-have-no-patience-for-a-slow-website (accessed on 21 March 2022).
4. Using Site Speed in Web Search Ranking. Available online: https://developers.google.com/search/blog/2010/04/using-site-speed-in-web-search-ranking (accessed on 21 March 2022).
5. Charnes, A.; Cooper, W.W.; Rhodes, E. Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* **1978**, *2*, 429–444. [CrossRef]
6. Daraio, C.; Kerstens, K.H.; Nepomuceno, T.C.C.; Sickles, R. Productivity and efficiency analysis software: An exploratory bibliographical survey of the options. *J. Econ. Surv.* **2019**, *33*, 85–100. [CrossRef]
7. Daraio, C.; Kerstens, K.; Nepomuceno, T.; Sickles, R.C. Empirical surveys of frontier applications: A meta-review. *Int. Trans. Oper. Res.* **2020**, *27*, 709–738. [CrossRef]
8. Nepomuceno, T.; Costa, A.; Daraio, C. Theoretical and empirical advances in the assessment of productive efficiency sincethe introduction of dea: A bibliometric analysis. *Int. J. Oper. Res.* **2023**, *46*, 505–549. [CrossRef]

9. Nepomuceno, T.; Santiago, K.; Daraio, C.; Costa, A. Exogenous crimes and the assessment of public safety efficiency and effectiveness. *Ann. Oper. Res.* **2020**, *316*, 1349–1382. [CrossRef]
10. Nepomuceno, K.; Nepomuceno, T.; Sadok, D. Measuring the internet technical efficiency: A ranking for the world wide web pages. *IEEE Lat. Am. Trans.* **2020**, *18*, 1119–1125. [CrossRef]
11. Saraiva, N.; Soares, A. A performance evaluation programmable architecture for choice of switching paradigm in hybrid optical networks. *IEEE Lat. Am. Trans.* **2016**, *14*, 4567–4572. [CrossRef]
12. Du, Y.; Li, C.; Hu, Q.; Li, X.; Chen, X. Ranking webpages using a path trust knowledge graph. *Neurocomputing* **2017**, *269*, 58–72. [CrossRef]
13. Nepomuceno, K.; de Oliveira, I.N.; Aschoff, R.R.; Bezerra, D.; Ito, M.S.; Melo, W.; Sadok, D.; Szabo, G. Quic and tcp: A performance evaluation. In Proceedings of the 2018 IEEE Symposium on Computers and Communications (ISCC), Natal, Brazil, 25–28 June 2018; pp. 00045–00051. [CrossRef]
14. Machado, L.; de Mello, J.; Roboredo, M. Efficiency evaluation of brazilian electrical distributors using dea game and cluster analysis. *IEEE Lat. Am. Trans.* **2016**, *14*, 4499–4505. [CrossRef]
15. Fenerich, F.; da Costa, S.; de Lima, E. Energy efficiency in industrial environments: Overview and research agenda. *IEEE Lat. Am. Trans.* **2017**, *15*, 415–422. [CrossRef]
16. Constant, R.; de Mello, J. Brazilian electricity distributors efficiency index based on non radial efficiency. *IEEE Lat. Am. Trans.* **2017**, *15*, 1657–1663. [CrossRef]
17. Nepomuceno, T.C.C.; Costa, A.P.C.S. Resource allocation with Time Series DEA applied to Brazilian Federal Saving banks. *Econ. Bull.* **2019**, *39*, 1384–1392.
18. Nepomuceno, T.C.C.; Daraio, C.; Costa, A.P.C.S. Combining multi-criteria and directional distances to decompose non- compensatory measures of sustainable banking efficiency. *Appl. Econ. Lett.* **2020**, *27*, 329–334. [CrossRef]
19. Serrano-Cinca, C.; Fuertes-Calle, Y.; Mar-Molinero, C. Measuring dea efficiency in internet companies. *Decis. Support Syst.* **2005**, *38*, 557–573. [CrossRef]
20. Martínez-Núñez, M.; Pérez-Aguiar, W.S. Efficiency analysis of information technology and online social networks management: An integrated dea-model assessment. *Inf. Manag.* **2014**, *51*, 712–725. [CrossRef]
21. Li, C.; Cheng, B.; Chen, J.; Gu, P.; Deng, N.; Li, D. A web service performance evaluation approach based on users experience. In Proceedings of the 2011 IEEE International Conference on Web Services, Washington, DC, USA, 4–9 July 2011; pp. 734–735. [CrossRef]
22. Klumpp, M.; Loske, D. Order picking and e-commerce: Introducing non-parametric efficiency measurement for sustainable retail logistics. *J. Theor. Appl. Electron. Commer. Res.* **2021**, *16*, 846–858. [CrossRef]
23. Ghosh, P.; Rau-Chaplin, A. Performance of dynamic web page generation for database-driven web sites. In Proceedings of the International Conference on Next Generation Web Services Practices, Seoul, Republic of Korea, 25–28 September 2006; pp. 56–63. [CrossRef]
24. Butkiewicz, M.; Madhyastha, H.V.; Sekar, V. Characterizing web page complexity and its impact. *IEEE/ACM Trans. Netw.* **2014**, *22*, 943–956. [CrossRef]
25. Najadat, H.; Al-Badarneh, A.; Al-Huthaifi, R.; Abo-Zaitoon, A.; Al-Omary, Y. Evaluating jordanian universities' websites based on data envelopment analysis. In Proceedings of the 2017 8th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 4–6 April 2017; pp. 159–164. [CrossRef]
26. Li, Z.; Li, J.; Li, X. A study on website operation efficiency evaluation: Based on dea model. In Proceedings of the 2009 International Conference on Management and Service Science, Beijing, China, 20–22 September 2009; pp. 1–6. [CrossRef]
27. Widmer, J.; Denda, R.; Mauve, M. A survey on tcp-friendly congestion control. *IEEE Netw.* **2001**, *15*, 28–37. [CrossRef]
28. Lychev, R.; Jero, S.; Boldyreva, A.; Nita-Rotaru, C. How secure and quick is quic? Provable security and performance analyses. In Proceedings of the 2015 IEEE Symposium on Security and Privacy, San Jose, CA, USA, 17–21 May 2015; pp. 214–231. [CrossRef]
29. Carlucci, G.; De Cicco, L.; Mascolo, S. Http over udp: An experimental investigation of quic. In SAC '15: Proceedings of the 30th Annual ACM Symposium on Applied Computing, SAC 2015: Symposium on Applied Computing, Salamanca, Spain, 13–17 April 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 609–614. [CrossRef]
30. Megyesi, P.; Krämer, Z.; Molnár, S. How quick is quic? In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 22–27 May 2016; pp. 1–6. [CrossRef]
31. Biswal, P.; Gnawali, O. Does quic make the web faster? In Proceedings of the 2016 IEEE Global Communications Conference (GLOBECOM), Washington, DC, USA, 4–8 December 2016; pp. 1–6. [CrossRef]
32. Reddy, N.R.; Reddy, P.C.; Padmavathamma, M. Efficient traffic engineering strategies for improving the performance of tcp friendly rate control protocol. *Futur. Internet* **2017**, *9*, 74. [CrossRef]
33. Kakhki, A.M.; Jero, S.; Choffnes, D.; Nita-Rotaru, C.; Mislove, A. Taking a long look at quic: An approach for rigorous evaluation of rapidly evolving transport protocols. In *IMC '17: Proceedings of the 2017 Internet Measurement Conference, IMC '17: Internet Measurement Conference, London, UK, 1–3 November 2017*; Association for Computing Machinery: New York, NY, USA, 2017; pp. 290–303. [CrossRef]
34. Peng, G.-Q.; Xue, G.; Chen, Y.-C. Network measurement and performance analysis at server side. *Futur. Internet* **2018**, *10*, 67. [CrossRef]

35. Jaloudi, S. Communication protocols of an industrial internet of things environment: A comparative study. *Futur. Internet* **2019**, *11*, 66. [CrossRef]
36. Olmedo, G.; Lara-Cueva, R.; Martínez, D.; de Almeida, C. Performance analysis of a novel tcp protocol algorithm adapted to wireless networks. *Futur. Internet* **2020**, *12*, 101. [CrossRef]
37. The Top 500 Sites on the Web. Available online: https://www.alexa.com/topsites/ (accessed on 21 March 2022).
38. Charles, V.; Aparicio, J.; Zhu, J. The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis. *Eur. J. Oper. Res.* **2019**, *279*, 929–940. [CrossRef]
39. Google Bigquery. Available online: https://cloud.google.com/bigquery/ (accessed on 21 March 2022).
40. Netravali, R.; Sivaraman, A.; Das, S.; Goyal, A.; Winstein, K.; Mickens, J.; Balakrishnan, H. Mahimahi: Accurate record-and-replay for http. In Proceedings of the USENIX ATC, Santa Clara, CA, USA, 8–10 July 2015; pp. 417–429.
41. Wang, X.S.; Balasubramanian, A.; Krishnamurthy, A.; Wetherall, D. How speedy is SPDY? In Proceedings of the 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14), Seattle, WA, USA, 2–4 April 2014; pp. 387–399.
42. Shephard, R.W. *Theory of Cost and Production Functions*; Princeton University Press: Princeton, NJ, USA, 1970.
43. Fare, R.; Lovell, C.K. Measuring the technical efficiency of production. *J. Econ. Theory* **1978**, *19*, 150–162. [CrossRef]
44. Cook, W.D.; Seiford, L.M. Data envelopment analysis (dea)—Thirty years on. *Eur. J. Oper. Res.* **2009**, *192*, 1–17. [CrossRef]
45. Charnes, A.; Cooper, W.; Golany, B.; Seiford, L.; Stutz, J. Foundations of data envelopment analysis for pareto-koopmans efficient empirical production functions. *J. Econom.* **1985**, *30*, 91–107. [CrossRef]
46. Nepomuceno, T.C.C.; Daraio, C.; Costa, A.P.C.S. Multicriteria ranking for the efficient and effective assessment of police departments. *Sustainability* **2021**, *13*, 4251. [CrossRef]
47. Daraio, C.; Simar, L.; Wilson, P.W. Central limit theorems for conditional efficiency measures and tests of the 'separability' condition in non-parametric, two-stage models of production. *Econom. J.* **2018**, *21*, 170–191. [CrossRef]
48. Badin, L.; Daraio, C. Explaining Efficiency in Non-parametric Frontier Models: Recent Developments in Statistical Inference. In *Exploring Research Frontiers in Contemporary Statistics and Econometrics*; Physica-Verlag: Heidelberg, Germany, 2012; Chapter 1; pp. 151–175. [CrossRef]
49. Nepomuceno, T.C.C. Data-driven Analytics for Socioeconomic Challenges in a Contemporary World. *Socioecon. Anal.* **2023**, *1*, 1–4.
50. Boezi, G.; Cottarelli, A.; Manimendra, I.; Pagmozzi, D. The Efficiency of European Universities and the Role of Exogenous Factors. *Socioecon. Anal.* **2023**, *1*, 103–125.