mathematics

MDPI

Article

Assessing the Performance of Nonprofit Organizations: A Regression-Based Absolute Weight-Restricted Benefit-of-the-Doubt Data Envelopment Analysis Model

Hemantha Herath [1] **and Don Galagedera** [2,*]

[1] Department of Accounting, Brock University, St. Catharines, ON L2S 3A1, Canada; hherath@brocku.ca
[2] Department of Econometrics and Business Statistics, Monash University, Melbourne, VIC 3145, Australia
* Correspondence: tissa.galagedera@monash.edu

Abstract

Performance evaluation becomes more complex when a nonprofit organization's program expense ratio (PER) is represented by alternative measures that capture the same underlying construct. There are differences in PER measurement procedures and therefore it is difficult to determine which single measure may provide the most appropriate basis for performance appraisal. Moreover, conventional aggregation methods may not adequately account for differences among these estimates. This study proposes a benefit-of-the-doubt (BoD) data envelopment analysis (DEA) framework to develop a composite measure of PER incorporating alternative estimates of PER. The proposed approach allows the relative importance assigned to each PER estimate to vary across years while imposing weight restrictions to prevent the DEA model from assigning excessively flexible or zero weights to individual PER estimates. The empirical application considers three alternative estimates of PER over 19 years. The proposed model ensures that all estimates contribute to the evaluation while retaining the BoD principle. The analysis further compares the resulting efficiency scores and rankings with those obtained from the BoD DEA model and examines the robustness of the results under alternative weight restrictions. The findings demonstrate how weight-restricted BoD DEA can provide a more balanced and transparent framework for aggregating alternative estimates of PER while allowing differences in their relative importance across years. The study contributes a structured approach to evaluating alternative measurements of the same performance construct, offering greater transparency in weight determination and a basis for more balanced performance assessment.

Keywords: performance evaluation; non-profit organization; efficiency measurement; data envelopment analysis; benefit-of-the-doubt model; weight restriction

MSC: 90C08

Check for updates

Academic Editors: Ke Fu and Jiayan Xu

Received: 23 September 2026
Revised: 22 October 2026
Accepted: 27 October 2026
Published: 29 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Nonprofit organizations (NPOs) in the United States have played an important role in delivering services that address social and economic needs and provide humanitarian support. As a result, these nonprofits are required to be accountable and transparent. Nonprofit performance is often measured by financial expense ratios. More to the point, donors, charity rating organizations and regulators are keen on knowing how donations are used in conducting mission-driven program activities. Consequently, the issue of how consistently an NPO uses donor funds to deliver program activities is an ongoing

research question that has received considerable attention in the accounting literature [1]. Program expense ratio (PER) is one of the frequently used performance indicators [1–3]. The proportion of a nonprofit organization's expenditure that is devoted to providing mission-centered program activities is measured by PER.

A higher PER is perceived to be better, as it indicates that a larger proportion of the total expenses is spent on delivering benefits to recipients [4]. High PER reporting nonprofit organizations are typically considered efficient. Charity watchdogs, donors and regulators commonly use PER as a primary metric to evaluate efficiency and accountability [5–8]. In this study, an analytical approach is proposed to evaluate whether a nonprofit organization has effectively allocated funds to program activities over time.

The program expense ratio can be measured in multiple ways due to differences in accounting treatment and reporting requirements [1,5,9]. As a result, multiple estimates of PER exist for the same NPO. In addition, differences in PER can arise due to the accounting treatment of joint costs. Prior research has identified that joint costs may arise when educational activities are combined with fund solicitation in mail campaigns. The common accounting issues are allocating fund raising expenses to program expenses and considering gifts-in-kind as revenue [10,11]. Existence of multiple PERs creates uncertainty regarding how an NPO's performance can be measured. More to the point, different PERs will result in different conclusions about the same organization's financial performance and transparency.

Charity Watch provides ratings for nonprofit organizations. A charity labeled NPO-X for anonymity has been operating in the United States for approximately twenty years. In 2014, as per its audited financial statements, NPO-X reported that 77 percent of its total expenditure was spent on mission-centered program activities. The remaining 23 percent was spent on fundraising activities and administration expenses. A more cautious approach used by the rating agency Charity Watch found that in the same year, only 50 percent of expenditure was used for mission-driven program activities and consequently designated NPO-X with a lower performance rating. Charity Watch uses a conservative approach to measure the program expense ratio. As per Charity Watch methodology, the program expenses are restated by reclassifying solicitation costs allocated to program activities as fundraising expenses. Consequently, the program expenses ratio is lower but more relevant and aligned with donor expectations of transparency. The conservative adjustments by Charity Watch are based on publicly available Internal Revenue Services (IRS) Form 990 joint cost disclosures and audited financial statements.

The NPO-X example illustrates how differences in the sources and preparation of financial information can lead to varying assessments of nonprofit performance. These discrepancies may arise from differences in expense classification, allocation of joint costs and other accounting adjustments. This highlights the limitation of relying on a single metric to assess performance and underscores the need for a robust methodology to evaluate charities' program spending.

The methodology proposed in this article incorporates three program expense ratios derived from the audited financial statements, IRS Form 990 and Charity Watch's restated values. Despite differences in accounting treatments and reporting practices, considering alternative PER measures collectively may provide a more holistic evaluation framework to assess whether an NPO has consistently allocated adequate donor funds to mission-driven program activities over time.

This article uses data envelopment analysis (DEA) to incorporate alternative program expense ratios for assessing NPO performance. DEA is a non-parametric technique widely used to evaluate the relative performance of homogeneous units. DEA provides a systematic approach for combining multiple performance indicators into a composite

measure. Traditional DEA models have limitations. Therefore, in this study, we propose a weight-restricted (WR) benefit-of-the-doubt (BoD) DEA model. The BoD approach offers a flexible and robust methodology for combining multiple performance indicators into a composite index [12]. BoD DEA model combines multiple estimates of PER into a composite estimate and enables performance comparison across years. Consequently, a more robust evaluation framework is proposed by incorporating three PER measures using the WR BoD DEA model. A few studies have used DEA models to evaluate NPO performance. However, integration of multiple PER estimates with weight-restriction has not received much attention in the nonprofit accounting literature. Therefore, evaluation of NPOs using the proposed WR BoD DEA model is a contribution to nonprofit accountability literature. Furthermore, the proposed approach also provides a practical tool for charity evaluators and donors to assess whether donor funds are used for mission related activities through sustained program funding.

Two primary research questions addressed in this article are: First, rather than influencing program expense ratios through joint cost allocation practices, do NPOs consistently allocate majority of the donor funds to mission-driven program activities? Second, instead of using a single PER measure, does the WR BoD DEA model, which simultaneously combines three PER estimates, provide a reasonable assessment of nonprofit organization performance? This article attempts to contribute to the literature on nonprofit organization accountability by proposing a WR BoD DEA model to address these questions.

The remainder of the paper contains six sections. A literature review on DEA applications in NPO performance evaluation is provided in Section 2. Section 3 describes the data. Section 4 presents the DEA methodology and model formulation. Section 5 discusses empirical results, followed by robustness check in Section 6. Section 7 concludes the study, highlighting limitations and suggesting directions for future research.

2. Literature Review

2.1. DEA Applications in Nonprofit Organizations

Data envelopment analysis (DEA) has been widely applied to evaluate organizational efficiency across diverse sectors, including healthcare (e.g., hospitals), education (e.g., schools and university departments), banking and finance (e.g., investment funds and insurance companies), manufacturing (e.g., production plants and supply chains), agriculture (e.g., farms and fisheries), energy and utilities (e.g., electricity companies and power stations), and the public sector (e.g., police departments and tax offices) [13–16]. However, its application to nonprofit organizations (NPOs), including charities, remains comparatively limited.

Early research established DEA as a promising alternative to conventional approaches for evaluating and monitoring nonprofit performance. DEA was proposed as a means of addressing the limitations of ad hoc and potentially arbitrary metrics used by regulators and policymakers to assess nonprofit outputs [17]. A critical review of [17], emphasizing the advantage of using DEA as DEA-based efficiency scores are unit-invariant, was suggested in [18]. Subsequent studies have applied DEA to assess nonprofit efficiency across a range of service contexts. A large-scale DEA model to evaluate more than one thousand field operating units of a nonprofit humanitarian organization involved in disaster relief, emergency communication and life-saving skills training was proposed in [19].

Fundraising efficiency has also received considerable attention. Application of DEA to U.S. symphony orchestras over a two-year period to assess the relative efficiency of converting fundraising expenditure into developmental spending has been suggested [20]. Regression analysis is subsequently used to investigate factors associated with fundraising inefficiency. Similarly, Ref. [21] evaluates British charities using administration and

fundraising expenses as inputs and voluntary income as an output. A two-stage DEA model that distinguishes fundraising efficiency from the efficiency of utilizing funds for cause-related activities is developed in [22].

Other studies extend DEA applications to specific nonprofit subsectors and examine the organizational factors associated with efficiency. Using data from Portuguese sports clubs, Ref. [15] investigates the efficiency of nonprofit sports clubs using DEA and examine how stakeholders contribute to differences in efficiency. A DEA model for nonprofit sports clubs in Portugal is proposed in [23], where the analysis with survey evidence is used to investigate the relationships between efficiency and stakeholder relationships. In a study of the relationship between government funding and the efficiency of nonprofit colleges using DEA, Ref. [24] finds that increases in government funding do not significantly influence efficiency on average.

Some studies consider program delivery efficiency and fundraising efficiency as two separate functions. For example, a study distinguishing Korean nonprofit organizations that are efficient in both program delivery and fundraising efficiency from those that are efficient in only one of these activities is developed in [25]. Similarly, Ref. [26] demonstrates the usefulness of DEA for benchmarking nonprofit organizations in the transportation sector and Ref. [16] incorporates beneficiary information into a conventional DEA model to assess relative nonprofit efficiency of 117 Irish charities.

Collectively, these studies demonstrate the usefulness of DEA for evaluating and benchmarking nonprofit organizations across different operational settings. They also illustrate the diversity of DEA applications, ranging from single-stage efficiency assessment to multistage models that distinguish fundraising, resource utilization, service delivery and organizational effectiveness.

Application of DEA to accounting measures, however, warrants additional considerations. Based on a review of 280 studies encompassing 322 DEA models that use accounting measures, Ref. [27] shows that accounting-based DEA models capture different organizational performance constructs and hence require careful selection and interpretation of variables. In the context of nonprofit organizations, these distinctions are even more important because financial ratios and existing performance evaluation systems may capture specific dimensions of organizational performance without necessarily providing a unified assessment across multiple indicators. Our study focuses only on one aspect of performance, specifically, alternative measures of the same construct, PER, rather than three unrelated dimensions of organizational performance and therefore does not suffer from problems usually encountered in variable selection in DEA.

2.2. DEA Model Specifications in Nonprofit Performance Evaluation

The nonprofit DEA literature encompasses several model structures designed to address different dimensions of organizational performance. Conventional DEA models are used to estimate relative efficiency while extensions such as two-stage, multistage and network DEA models capture interrelated operational activities and intermediate outputs. For example, Ref. [22] develops a linked two-stage model where funds raised constitute an intermediate product linking the two stages, allowing the analysis to distinguish the efficiency of obtaining resources from the efficiency of deploying them. This approach is extended by proposing a four-stage DEA model incorporating fundraising, capacity, quality and effectiveness into a network framework in [14]. These models recognize that nonprofit performance is multidimensional and that efficiency in one activity does not necessarily imply efficiency throughout the organization. Other studies use different model specifications to investigate scale effects and organizational characteristics. Several studies compare efficiency estimates under constant returns to scale and variable returns to scale,

thereby examining whether differences in performance are associated with technical or scale efficiency [15,25].

Despite these methodological variations, the primary focus of existing studies is generally on assessing the relative efficiency of organizations, activities or operating units. Some studies incorporate financial information from IRS Form 990 or other accounting sources, but the alternative estimates of a particular financial performance measure are not themselves the principal objects of evaluation.

2.3. Financial Ratios and Nonprofit Performance Assessment

Financial ratios, particularly the program expense ratio (PER) and overhead ratios, are widely used to evaluate nonprofit performance. These measures provide accessible indicators of how organizational resources are allocated between program activities, administration and fundraising. Reviews of nonprofit program-ratio management in [1,28] document the continued reliance on financial indicators in nonprofit performance assessment. Accounting-based measures also remain common in empirical research examining program spending and overhead-management practices [29].

However, financial ratios provide only a partial view of nonprofit efficiency because they do not capture the full range of organizational inputs, outputs and outcomes. This limitation is highlighted by a study that combines operational and financial data to evaluate and rank Habitat for Humanity affiliates [30]. The study compares rankings based on overhead and administrative expense ratios with those derived from DEA and stochastic frontier analysis. These findings illustrate that accounting-based ratios and frontier-based efficiency measures may provide different assessments of organizational performance. Although Ref. [31] concerns financial institutions rather than charities, it illustrates the complementary roles of accounting indicators and frontier-based methods in performance assessment.

Although DEA simultaneously considers multiple inputs and outputs, conventional ratio analysis typically examines individual output-to-input (financial ratios, each of which provides information about a specific dimension of performance). However, the two measures can complement each other when used jointly. Such complementarity is illustrated by [32], which compares DEA and ratio analysis in assessing the performance of English District Health Authorities in the provision of perinatal care. Their study examined the extent to which the two methods agreed in their performance assessments and found that their results could agree reasonably closely depending on the indicators used and how they were combined into an overall performance measure. In an application to the oil and gas industry, Ref. [33] investigated the relationship between DEA and traditional accounting ratios as measures of firm performance. They demonstrated that DEA could provide additional information and complement conventional financial ratio analysis. Nevertheless, the two approaches have the potential to produce different performance assessments because they aggregate and evaluate performance differently. Ref. [34] developed a mathematical framework connecting DEA efficiency scores with conventional output–input ratios, establishing a theoretical basis for understanding the relationship between the two approaches.

Overall, DEA evaluates relative efficiency by considering multiple inputs and outputs within a common analytical framework. On the other hand, individual ratios reduce performance to a single-dimensional measure. Therefore, a given ratio measure may overlook other relevant dimensions. However, combining DEA with ratio analysis can provide a more comprehensive evaluation of organizational or firm performance. Newer methodological work explicitly combines DEA with ratio analysis. Accordingly, using DEA alongside conventional ratio analysis can provide a more comprehensive assessment of organizational or firm performance [35].

Interpretation of nonprofit financial ratios is further complicated by differences in data sources and accounting practices. Rating agencies such as CharityWatch, Charity Navigator, and the BBB Wise Giving Alliance use financial information from audited financial statements, IRS Form 990 filings and other assessment criteria to inform donors, regulators, and policymakers. However, differences in reporting sources, expense classifications, joint-cost allocations and accounting adjustments may produce different estimates of the same financial ratio. Discrepancies between information reported in nonprofit annual reports and IRS Form 990 filings have been documented [36]. Charity Navigator relies primarily on Form 990 data supplemented by other assessment criteria [22]. CharityWatch draws information from audited financial statements, Form 990 filings and accounting adjustments. Consequently, alternative PER estimates may produce different assessments of the same organization's program spending even when they are intended to measure the same underlying construct. This creates a methodological challenge: relying on a single estimate may not adequately reflect the information available from alternative reporting sources while treating all estimates as equally reliable may also be inappropriate.

2.4. Research Gap

To the best of our knowledge, the literature reviewed here has not examined how a DEA-based framework can be used to evaluate and aggregate alternative estimates of the same underlying performance construct. This issue is particularly relevant to PER because estimates derived from different reporting sources and accounting procedures may yield divergent assessments of nonprofit organization performance. To address this gap, this study proposes a DEA-based methodology to estimate a composite score incorporating three alternative estimates of PER. The framework allows relative importance assigned to each PER estimate to vary across years while imposing empirically derived weight restrictions to limit excessive weight flexibility. In doing so, the study extends the application of DEA to systematically evaluate alternative measurements of the same performance construct.

3. Data

To demonstrate application of the proposed WR BoD DEA model (see Section 4 for details), publicly available financial information obtained from a nonprofit organization is used. The nonprofit organization is referred to as NPO-X to preserve organizational anonymity. Tax-exempt organizations in the United States are required to file IRS Form 990, 990-EZ or 990N. IRS Form 990 is publicly available and provides detailed disclosures regarding an organization's financial position, functional expenses, governance, and operational activities. NPO-X has operated as a registered charitable organization and publishes both audited financial statements and IRS Form 990 filings on its website. The availability of both reporting sources enables a detailed examination of alternative PER estimates based on different accounting treatments.

IRS Form 990 provides useful information to evaluate program expenditure. The statement of functional expenses (Part IX) lists total expenses, program-service expenses, management and general expenses, and fundraising expenses. Line 26 titled "Joint cost" is required to be completed if the organization reported joint costs from a combined educational campaign and fundraising solicitation. Part XII provides a reconciliation of expenses reported in audited financial statements with expenses reported on the Form 990 return. Items include total expenses as reported in audited financial statements adjusted for donated services, other losses, prior-year adjustments, investment expenses not included in Form 990 and other reporting differences. Information provided in (Part XIII) of Form 990 discloses supporting schedules regarding functional expenses.

The Financial Accounting Standards Board Accounting Standards which deal with nonprofit organizations are covered by FASB ASC 958-720. It provides guidance on how NPOs classify, recognize, and report operational costs and report functional expense allocations. The current AICPA Statement of Position (SOP) 98-2, *Accounting for Costs of Activities of Not-for-Profit Organizations and State and Local Governmental Entities that Include Fundraising* came into effect in 1998 and provides guidance on joint cost allocation. Based on this guidance, a nonprofit is allowed to allocate a portion of the joint costs associated with a combined fundraising solicitation and an education campaign. These combined expenses may include the cost of direct mailing campaign, educational leaflets, postage costs, etc. The expenditure must meet the criteria of purpose, audience and content. Although joint cost allocation is allowed when performed in a rational, systematic and consistent manner by a nonprofit organization, there is doubt as to whether such joint costs should be treated as program expenses especially when an NPO tends to manage program expense ratios [1].

The CharityWatch efficiency ratio calculation criteria for restating the audited financial statements when joint costs are reported are used in this article. More specifically, when computing program expense ratios, Charity Watch adjusts solicitation expenses out of the reported program expenses and adds them to fundraising expenses. This restatement of joint costs prior to computing a charity's efficiency ratios and letter grade ratings is a conservative approach. Nonetheless, this conservative approach is perhaps what donors reasonably expect, since often the solicitation of donor funds using a mailing campaign with some program elements in its description may not be construed as a mission-driven program activity which directly benefits the charity recipients. Since joint costs are removed from program expenses and reallocated to fundraising, the program expenses ratios computed by Charity Watch for rating are lower than those reported in audited financial statements.

The data covering the period 2006–2024 were obtained from the audited financial statements and IRS Form 990 filings available on the NPO-X website. The period from 2006 to 2024 was selected because it encompassed both favorable and unfavorable agency ratings. NPO-X was founded in 2003 and became an independent NPO in 2005. NPO-X expanded rapidly from 2006 to 2014. In 2016, NPO-X was given a lower rating by rating agencies, including CharityWatch, because rating agencies believed that NPO-X accounting and spending reflected poor efficiency and there were governance challenges [10]. Governance reforms were introduced in 2017 to 2020, resulting in reversal of lower ratings in subsequent years.

The functional expenses are disclosed in the audited financial statements. In addition, they are also reported in Part IX, the statement of functional expenses in IRS Form 990, along with disclosures which may include donated services and use of facilities, investment expenses and other expenses, which enable the reconciliation of expenses as per the audited financial statements and Form 990. The reconciliation of total expenses reported in the audited financial statements with those reported in IRS Form 990 for NPO-X in 2017 is shown in Table 1.

Table 1. Reconciliation of expenses.

Reconciling Item	Amount (in $)
Expenses per audited financial statements	270,211,475
Less: Donated services and use of facilities	(39,027,136)
Add: Investment expenses not included in Form 990	643,006
Total expenses reported on Form 990	231,827,345

Reconciliation of the expenses reported in Table 1 as per the two statements indicates that the total expenses reported in the audited financial statements and Form 990 are different. This difference may result in program expense ratios which are quite divergent. This article computes three program expense ratios. One of the program ratios is computed using the expense information provided in the audited financial statements. Another program ratio is computed using the expense information provided in Form 990. To compute the restated program expense ratio, Charity Watch conservative adjustment procedure is followed. Specifically, the program expenses reported in the audited financial statements are restated by subtracting the joint fundraising expenses allocated to program expenses and reclassifying them as fundraising expenses. These joint costs associated with the combined fundraising, and educational campaigns are obtained from line 26 of Part IX. Table 2 presents the restated program expenses using Charity Watch conservative methodology for 2017. Three estimates of PER, defined as program expenses divided by total expenses, are constructed using the audited financial statements, IRS Form 990 filings, and restated expenses. The resulting PER estimates for the period 2006–2024 are presented in Table 3.

Table 2. Reported and restated expenses of NPO-X for year 2017.

	Program Expenses	**Management and General Expenses**	**Fundraising Expenses**	**Total Expenses**
Form 990	165,835,429	12,981,666	53,010,250	231,827,345
Joint cost reported in Form 990	16,339,783			
Adjust to fundraising expenses			16,339,783	
Restated expenses	149,495,646	12,981,666	69,350,033	231,827,345
Audited financial statements	203,645,836	13,278,844	53,286,795	270,211,475

Table 3. Input–output data for DEA.

Year	**Input (x)**	**Different PER Estimates**		
		Form 990	**Audited**	**Restated**
		Output y_1	**Output y_2**	**Output y_3**
2006	1	0.7565	0.7565	0.6686
2007	1	0.7852	0.7870	0.7411
2008	1	0.6416	0.8252	0.5621
2009	1	0.6170	0.8132	0.4803
2010	1	0.6492	0.8199	0.4547
2011	1	0.6788	0.8278	0.5503
2012	1	0.7287	0.8159	0.5799
2013	1	0.7414	0.7992	0.5771
2014	1	0.7643	0.8064	0.5993
2015	1	0.7463	0.7751	0.6123
2016	1	0.7050	0.7551	0.6072
2017	1	0.7153	0.7537	0.6449
2018	1	0.7204	0.7753	0.6400
2019	1	0.7113	0.7827	0.6294
2020	1	0.7116	0.7938	0.6392
2021	1	0.7105	0.7558	0.6314
2022	1	0.7166	0.7515	0.6318
2023	1	0.7054	0.7582	0.6257
2024	1	0.7018	0.7634	0.6182

Figure 1 graphs the NPO-X PER estimates over the investigation period. The three plotted series exhibit similar broad trends over time but differ markedly in their levels.

Overall, the audited PER consistently exceeds the Form 990-based PER and restated PER throughout the study period. It appears that audited financial statements report a higher proportion of expenditure allocated to program activities. All three measures increase between 2006 and 2007. Form 990-based PER and especially the restated PER decline sharply during 2008–2010. The restated PER reached its lowest value in 2010 before recovering steadily over subsequent years. In contrast, the audited PER remains comparatively stable throughout the study period fluctuating between approximately 0.75 and 0.83. After 2011, the differences between the three measures become more stable. Although the three series move in the same direction over time, the audited estimates consistently remain the highest and the restated estimates the lowest, suggesting systematic differences between the reporting methodologies rather than random variation.

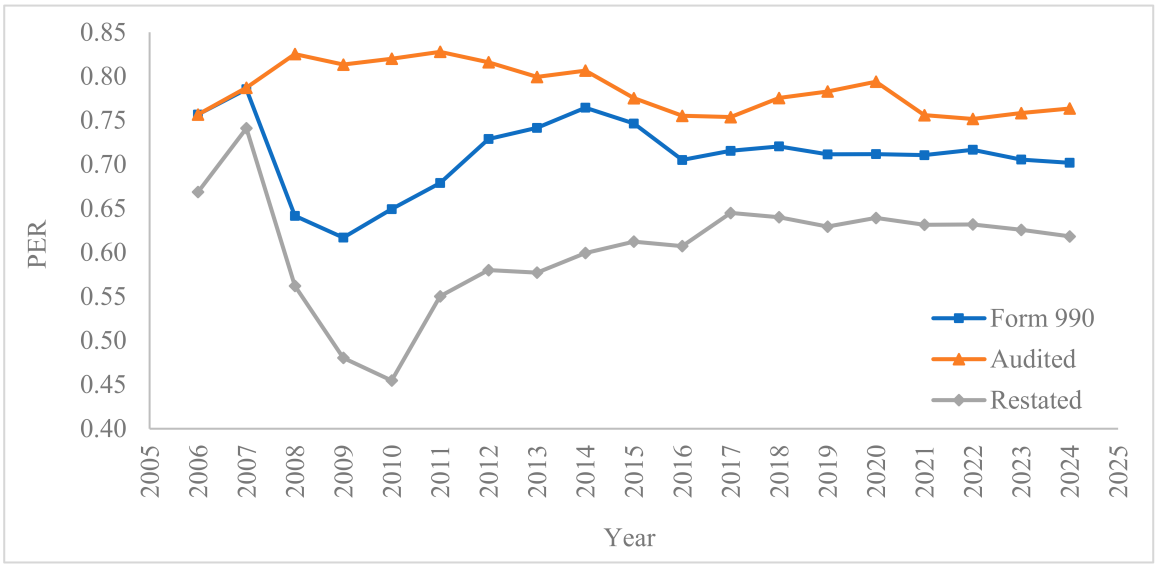


Figure 1. Different PER estimates from 2006 to 2024.

The observed differences among the three PER estimates provide the motivation for the DEA. Rather than relying on a single PER measure, the WR BoD DEA model evaluates each year using all three estimates simultaneously allowing a composite assessment that reflects the information contained in each reporting approach while accounting for their systematic differences.

4. Method

4.1. DEA-Based Performance Appraisal

DEA is a non-parametric mathematical programming technique that evaluates the relative efficiency of a set of homogeneous units referred to in the literature as decision-making units (DMUs) [37]. In relative efficiency appraisal, DEA model can accommodate multiple inputs and multiple outputs. DEA does not require the specification of a functional relationship between inputs and outputs [37,38]. This is considered an important advantage of using DEA over traditional parametric approaches. Further, parametric approaches require an assumption regarding the statistical distribution of inefficiency. Instead, DEA constructs a best-practice frontier from the observed data and measures the performance of each DMU under investigation relative to this best-practice frontier [38,39]. In the conventional DEA framework, each DMU is allowed to select a set of input and output weights that maximize its own efficiency score subject to the constraint that the same weights do not assign an efficiency score greater than one to any other DMU. Consequently, DMUs located on the efficient frontier receive an efficiency score of one and are considered

relatively efficient while those DMUs that lie below the frontier receive scores less than one. The DMUs that lie below the frontier are deemed inefficient, indicating that there is potential for performance improvement relative to the best performers.

The original DEA model proposed by [37], commonly known as the CCR model, assumes constant returns to scale (CRS) implying that outputs change proportionally with inputs. The CCR model was extended in [39] by incorporating variable returns to scale allowing efficiency to be separated into technical efficiency and scale efficiency. Depending on the objective of the study, DEA model may be specified as input-oriented where the objective is to minimize inputs while maintaining the current output levels or output-oriented where the objective is to maximize outputs given the existing input levels. The choice of orientation and returns to scale should reflect the operational characteristics and managerial objectives of the evaluated DMUs.

4.2. Empirical Model Formulation

Our empirical setting differs from the conventional DEA framework which typically evaluates the relative efficiency of multiple decision-making units (DMUs) using multiple inputs and outputs. Instead, we analyze a single nonprofit organization (NPO-X) over a 19-year period (2006–2024). For each year, three alternative estimates of the program expense ratio (PER) are available. These three measures represent alternative indicators of the same underlying performance construct.

The objective is to combine these three PER estimates into a single composite performance measure and evaluate the relative performance of NPO-X across the 19 years. To achieve this, each year is treated as a DMU, while the three PER estimates are specified as outputs. Because all observations relate to the same organization, we assume a common unit input for every year. Specifically, the input is normalized to one for all DMUs allowing differences in performance to be attributed solely to variations in the three PER measures. In this case, DEA is not used in the traditional economic production setting, where multiple inputs are transformed into outputs. Instead, a BoD DEA model is formulated to compare a combination of the three alternative estimates of the same underlying construct (PER) across years. As shown by [40,41], the BoD approach provides a DEA-based framework for constructing a composite index by assigning data-driven weights to the different dimensions. A similar application is available in [42]. The authors apply DEA to multiple performance indicators with a common unitary input to evaluate macroeconomic performance of OECD countries. Normalization of the inputs eliminates differences in input usage allowing the analysis to focus exclusively on the relative performance captured by different dimensions (outputs). We consider three alternative estimates of PER as outputs. These outputs are derived from (i) information reported in IRS Form 990 filings, denoted by y_1; (ii) audited financial statements, denoted by y_2; and (iii) the application of the restatement methodology, denoted by y_3. Although all three measures estimate the same underlying performance construct, they differ because of variations in reporting practices and measurement methodologies. Under this framework, relative PER performance of NPO-X in year k is computed using the output-oriented CRS DEA model presented in model (1).

Let, y_{rk} = output r of year k (r = 1, 2, 3; k = 1, 2, …, 19) > 0 and v_r = the weight associated with output y_{rk}. A BoD DEA model that assesses relative efficiency in year k may be written as

$$\text{Maximize } \sum_{r=1}^{3} v_r y_{rk} \tag{1}$$

Subject to

$$\sum_{r=1}^{3} v_r y_{rj} \leq 1, j = 1, 2, \ldots, 19$$

$$v_r \geq 0$$

Ref. [43] describes model (1) as a CCR DEA model in a pure-output setting with a dummy input normalized to one. The model constructs an efficient frontier comprising the years that exhibit the best combination of the three output measures, y_{rk}: r = 1, 2, 3. An efficiency score bounded between zero and one is obtained for each year. A score of one indicates that the year lies on the efficient frontier, whereas a score below one indicates that the year is dominated by at least one other year in terms of the combined PER measures. Because all years are evaluated using the same normalized input, the resulting efficiency scores reflect the relative performance of the combined alternative PER estimates rather than differences in resource utilization. Thus, model (1) provides a composite assessment of program expense performance based solely on the information contained in the three alternative PER measures.

Model (1) determines the weights v_r assigned to outputs without imposing subjective preferences. The efficiency score computed in model (1) for year k (labeled θ_k) is a relative efficiency estimate of composite PER measure for year k. $\theta_k = \sum_{r=1}^{3} v_r^* y_{rk}$ where $v_r^* : r = 1, 2, 3$ are the optimal weights obtained in a model (1) solution.

A distinguishing feature of model (1) is that each DMU (year) is permitted to select output weights that maximize its own relative efficiency score. This flexibility allows each year to be evaluated under its most favorable conditions without requiring prior specification of the relative importance of the outputs. However, model (1) may also produce weighting schemes that are inconsistent with managerial judgment or practical reality. Model (1) may assign extremely small or zero weights to variables that are regarded as important effectively ignoring their contribution to the performance evaluation. Such weight flexibility may result in efficiency estimates that are difficult to justify from a policy or operational perspective. Consequently, incorporating weight restrictions has become a common extension of conventional DEA to ensure that the resulting efficiency measures reflect both the observed data and relevant value judgments. When weight restrictions are imposed, discrimination between DMUs improves [44]. Weight flexibility nevertheless remains a critical issue in current BoD research [45].

There are various approaches for incorporating weight restrictions. The Assurance Region (AR) method proposed by [46] is one of the most widely adopted. Rather than specifying absolute values for individual weights, the AR approach restricts the ratios between selected pairs of weights to lie within predetermined lower and upper bounds. These bounds reflect prior knowledge, expert opinion, policy considerations, or empirical evidence regarding the relative importance of the variables. Another method of weight restriction is to impose constraints on virtual outputs $v_r y_{rk}$:r = 1, 2, 3. [47]. An investigation of weight restrictions in DEA is available in [48].

In this paper, we adopt the absolute weight restriction approach proposed by [48]. Conventional AR restrictions require bounds on relative weights, which may be difficult to specify when no sufficiently defensible external information is available. Our empirical approach instead uses regression-derived information from the three alternative PER indicators to inform the admissible weight set. This provides a data-informed basis for restricting weight flexibility while retaining the ability to construct a composite performance measure. Specifically, we regress the three output indicators, y_r (the PER measures obtained from Form 990, audited financial statements, and restated information), on the single input x, which is normalized to one.

First, we regress the three indicators y_{rk}: r = 1, 2, 3 on the single input x = 1. The regression equation we estimate is

$$x = \sum_{r=1}^{3} \beta_r y_r + \vartheta \quad (2)$$

In model (2), the intercept is set at zero. This restriction is appropriate because the purpose of the regression is not to estimate a conventional explanatory relationship, but rather to obtain relative measures of output importance for defining DEA weight restrictions. A zero intercept ensures that the estimated coefficients β_r: $r = 1, 2, 3$ represent the contribution of each output per unit of input. This is consistent with the DEA production framework, where zero input implies zero attainable output. Given that the input is normalized to one for all observations, the estimated coefficients provide numerical benchmarks for the relative weights assigned to the three alternative PER measures. The proposed formulation uses empirically estimated relationships to construct the admissible weight set.

Let $\hat{\beta}_r : r = 1, 2, 3$ denote the optimal parameter values obtained in the regression analysis. Then $\hat{\beta}_r$ can be interpreted as the resource units that NPO-X uses in a year on average per unit of y_r. We use $\hat{\beta}_r$ to set lower bounds on the weight associated with output y_r. Initially, we impose the restrictions $v_r \geq 0.75\hat{\beta}_r$, $r = 1, 2, 3$ and later in the robustness check we consider different multiples of $\hat{\beta}$. Ref. [45] discusses the issue of excessive weight flexibility in DEA and emphasizes the role of weight restrictions in improving benchmarking and the interpretability of efficiency scores. In the present study, the imposed weight restrictions ensure that all three outputs make a meaningful contribution to the efficiency assessment and prevent the model from assigning negligible or zero weights to some outputs in any given year to obtain artificially high composite PER relative efficiency scores. This restriction therefore reflects the judgment that all three outputs are considered sufficiently important and must be adequately represented in the evaluation.

Finally, we obtain an estimate of relative efficiency in year k by solving the following WR BoD DEA model.

$$\text{Maximize } \sum_{r=1}^{3} v_r y_{rk} \tag{3}$$

Subject to

$$\sum_{r=1}^{3} v_r y_{rj} \leq 1, j = 1, 2, \ldots, 19$$

$$v_r \geq 0.75\hat{\beta}_r,\ j = 1, 2, 3$$

The output-oriented interpretation of model (3) is discussed in detail by [49]. More generally, Ref. [50] provides an extensive review of the role and implementation of weight restrictions in BoD DEA applications.

4.3. Input–Output Data for DEA

The dataset presented in Table 3 contains three alternative estimates of PER, denoted by y_1, y_2 and y_3 calculated annually over the period 2006–2024. Each year is treated as a DMU in the DEA framework. Since the analysis focuses on evaluating the relative performance of the three PER estimates over time, a common input, normalized to unity and denoted by x is assigned to each DMU.

5. Results and Discussion

The regression coefficients obtained in model (2) solution are $\hat{\beta}_1 = 0.1239$, $\hat{\beta}_2 = 0.9601$ and $\hat{\beta}_3 = 0.2609$. Table 4 presents the WR BoD DEA results obtained in model (3). The adopted WR BoD DEA model selects the most favorable combination of the three PER estimates subject to the imposed weight restrictions. Therefore, the efficiency scores represent each year's overall PER performance rather than performance on any single measure. The results in Table 4 reveal consistently high levels of performance over the study period. The relative efficiency scores range from 0.9347 to 1 with an average of 0.9646. The narrow range of efficiency scores indicates that PER efficiency estimates are stable over time, with only minor year-to-year variation.

Table 4. Relative efficiency scores obtained in the WR BoD DEA model.

Year (DMU)	Efficiency	Rank
2006	0.9534	13
2007	1	1
2008	0.9896	5
2009	0.9594	11
2010	0.9640	9
2011	0.9934	2
2012	0.9920	3
2013	0.9760	7
2014	0.9897	4
2015	0.9594	12
2016	0.9347	19
2017	0.9416	15
2018	0.9626	10
2019	0.9671	8
2020	0.9800	6
2021	0.9406	17
2022	0.9370	18
2023	0.9414	16
2024	0.9448	14
Average	0.9646	

The year 2007 is efficient with efficiency score of 1. It establishes the best-practice frontier against which all other years are evaluated. The efficiency scores of the other years are below 1. However, all relative efficiency scores exceed 0.93 implying relatively small deviations from the frontier over the 19-year period. Although only one year attained full efficiency, many years are only marginally below the frontier. For example, 2008 (0.9896), 2011 (0.9934), 2012 (0.9920), 2014 (0.9897) and 2020 (0.9800) are all within two percentage points of the efficient frontier indicating very similar overall performance. These years may therefore be regarded as near-efficient observations. The lowest efficiency score was observed in 2016 (0.9347), followed by 2022 (0.9370) and 2021 (0.9406). Even in these years, efficiency remained above 93% suggesting that performance deteriorated only modestly relative to the benchmark. These years therefore represent the greatest potential for improvement according to the WR BoD DEA model.

Figure 2 provides a visual representation of relative efficiency scores. Despite some year-to-year fluctuations, the fitted trend line exhibits a negative slope (-0.0022), suggesting a gradual deterioration in relative efficiency over the study period. NPO-X had rapid expansion in the early years (2006–2014), during which funds were used for non-program activities such as conferences, employee events, and executive travel, resulting in lower adjusted PER (Form 990 and restated) [10]. Over the same period, audited financial statements recognized these expenses as program expenses (joint costs), resulting in the higher PERs. In 2016, this spending controversy came to light and NPO-X was rated lower by rating agencies based on the restated PER. Although reforms were proposed in subsequent years, as shown by Form 990 based PER and restated PER, joint cost allocation issues continued resulting in decreasing ratios over the years.

Since the WR BoD DEA model evaluates each year relative to the best-performing years in the sample, the observed decline does not necessarily imply an absolute deterioration in PER; rather, it indicates that later years are relatively less efficient compared with the historical efficiency frontier defined by the sample period. BoD computes relative efficiency scores. For example, BoD efficiency score in 2024 is 0.9448. This does not mean IPO-X has achieved approximately 94% of some theoretical maximum level of performance. Instead,

it indicates that, in year 2024, PER overall performance is 94% of the best-performing observation in the dataset. Thus, the scores provide a measure of relative performance rather than an absolute measure of efficiency. Hence, DEA rankings of DMUs based on relative efficiency scores is more appropriate.

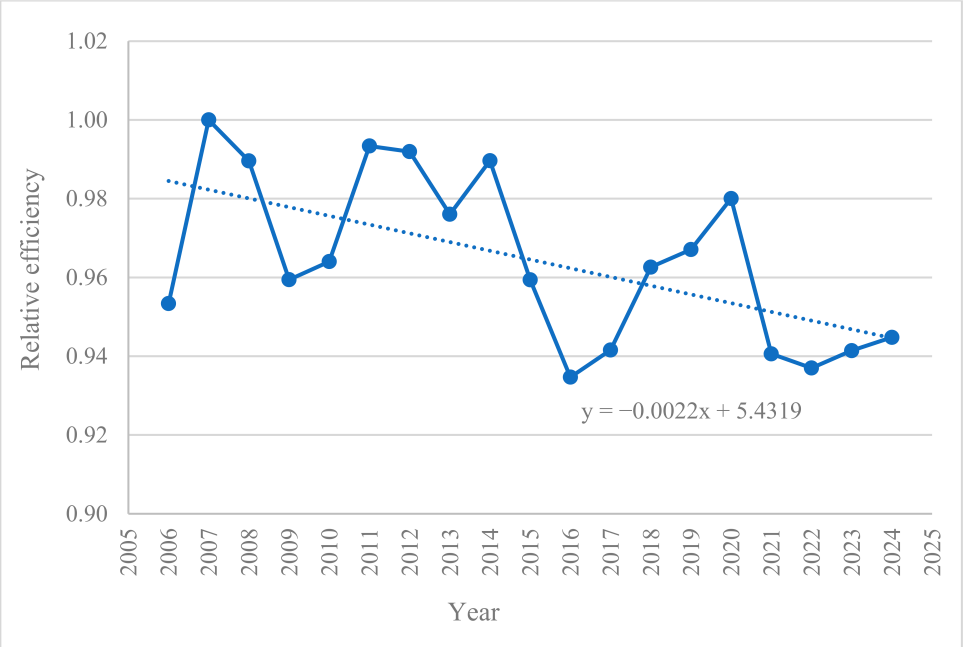


Figure 2. Graphical illustration of relative efficiency.

Column 3 in Table 4 presents the rankings associated with the relative efficiency scores. The results identify 2007 as the benchmark year with an efficiency score of 1 and a rank of one. No other year produces a superior combination of the three PER estimates under the AR restrictions imposed. The next highest-ranked years are 2011 (rank 2), 2012 (rank 3), 2014 (rank 4) and 2008 (rank 5). These years consistently lie close to the efficient frontier and may be regarded as representing periods of strong program expense performance relative to other years. The lowest-ranked years are 2016 (ranked 19), 2022 (ranked 18), 2012 (ranked 17) and 2023 (ranked 16). These years exhibit the lowest composite PER performance after simultaneously considering all three estimates. An interesting feature is that the highest rankings are concentrated in the earlier and middle parts of the study period (2007–2014) whereas most of the lowest-ranked years occur after 2016. This suggests a modest deterioration in relative program expense performance during the latter years of the sample. This should be interpreted as a decline in relative rather than absolute performance.

There are several practical and policy implications of using the proposed model. Ratings of charities are used by donors in their funding decisions. While financial ratios are still widely used, there has been continuous focus in both academic research and practice on improving charity ratings, so that donors can make more informed decisions. The proposed DEA model, which considers three diverse PERs and combines them into a single metric, provides a more balanced and reliable efficiency rating that would better inform donors, as is clearly indicated by the controversial year 2016 having the lowest DEA efficiency score. From an organizational perspective, using reliable, transparent and consistent efficiency ratios allows donors to favorably see the charity and continue to provide funds. In addition, there may be reputation effects whereby existing donors may encourage more donors, resulting in less uncertainty in fundraising. Finally, from a government policy perspective, reliable efficiency ratios may be valuable for assessing charity governance aspects.

The optimal output weights obtained in the WR BoD DEA model are presented in Table 5. The most striking feature of the results is that the optimal output weights are identical across the 19-year period. Except for 2006, every year receives the same weighting scheme. A desirable implication here is that differences in efficiency are driven primarily by differences in the PER estimates themselves rather than by changes in the optimal weights.

Table 5. Optimal output weights obtained in the WR BoD DEA model (model (3)).

Year (DMU)	v_1^*	v_2^*	v_3^*
2006	0.3672	0.7201	0.1957
2007	0.0929	0.9938	0.1957
2008	0.0929	0.9938	0.1957
2009	0.0929	0.9938	0.1957
2010	0.0929	0.9938	0.1957
2011	0.0929	0.9938	0.1957
2012	0.0929	0.9938	0.1957
2013	0.0929	0.9938	0.1957
2014	0.0929	0.9938	0.1957
2015	0.0929	0.9938	0.1957
2016	0.0929	0.9938	0.1957
2017	0.0929	0.9938	0.1957
2018	0.0929	0.9938	0.1957
2019	0.0929	0.9938	0.1957
2020	0.0929	0.9938	0.1957
2021	0.0929	0.9938	0.1957
2022	0.0929	0.9938	0.1957
2023	0.0929	0.9938	0.1957
2024	0.0929	0.9938	0.1957
Average	0.1073	0.9794	0.1957

The weights reported in Table 5 reveal a high degree of stability over the study period. The similarity of the weights across years indicates that the model consistently identifies a comparable weighting structure for combining the three alternative PER measures. This suggests that the relative contribution of the three PER measures to the composite PER efficiency score remains stable over time. Among the three alternative PER measures, the audited PER consistently receives the largest weight, followed by the restated PER and the Form 990 based PER. The consistent ranking of the weights also suggests that the audited PER played the dominant role in defining the composite performance measure under the WR BoD DEA optimization framework. The limited variation in the weights is also indicative of the effectiveness of the constraints imposed. DEA weights are scale-dependent and therefore the absolute magnitude of these weights may not be interpreted as an indicator of PER measure importance. For a meaningful interpretation of the weights, we normalize them. For example, in the year 2006, the optimal weights are 0.3672, 0.7201 and 0.1957. Their sum is 1.277. Then the normalized weights would be approximately 28.3%, 56.4% and 15.3%. These provide an intuitive description of the relative contribution of each PER measure. Hence, for ease of interpretation, the raw weights were normalized to sum to one. The average normalized weights indicate that the audited PER measure contributes approximately 76% of the composite PER measure, while restated PER and Form990 PER contribute approximately 15% and 8%, respectively.

6. Robustness Check

As a robustness check, efficiency scores obtained from the WR BoD DEA model (model (3)) are compared with those from the BoD DEA model (model (1)). The results show a high

degree of consistency between the two approaches. Both models identify similar periods of relatively high and low efficiency with 2007 representing the highest-performing year and 2016 representing the lowest-performing year. However, the WR BoD DEA model produces slightly lower efficiency estimates than the BoD DEA model. This difference is expected because the BoD DEA framework allows freedom to select the most favorable weight structure, whereas the WR BoD DEA approach evaluates performance through an internally derived composite measure under certain weight restrictions. Under the BoD DEA model, 2007, 2011, 2012 and 2014 are deemed DEA-efficient. On the other hand, WR BoD DEA model provides greater differentiation among high-performing years avoiding the occurrence of multiple efficiency scores of 1. The average normalized weights computed in the weight unrestricted BoD DEA model are $v_1^* = 0.2089$, $v_2^* = 0.7245$ and $v_3^* = 0.0667$. In this case, audited PER measure contributes approximately 72% of the composite PER measure, while restated PER and Form 990 based PER contribute approximately 21% and 7%, respectively.

The optimal weights obtained in the unrestricted BoD DEA model are presented in Table 6. In this case, the DEA appraisal assigns a zero weight to v_1 in six years, v_2 in two years, and v_3 in eight years. For example, in the DEA appraisal using BoD DEA model, the optimal solutions for 2009, 2010 and 2011 assign zero weights to v_1 and v_3. This is an indication that, when using BoD DEA model, two of the three outputs are effectively excluded from the DEA-based performance appraisal in 2009, 2010 and 2011. These observations justify using a weight-restricted BoD DEA model.

Table 6. Optimal output weights obtained in the BoD DEA model (model (1)).

Year (DMU)	v_1^*	v_2^*	v_3^*
2006	1.2735	0.0000	0.0000
2007	1.2735	0.0000	0.0000
2008	0.0000	1.0575	0.2264
2009	0.0000	1.2080	0.0000
2010	0.0000	1.2080	0.0000
2011	0.0000	1.2080	0.0000
2012	0.1658	0.9923	0.1199
2013	0.2634	0.9905	0.0000
2014	0.2634	0.9905	0.0000
2015	0.6134	0.6587	0.0000
2016	0.1658	0.9923	0.1199
2017	0.1658	0.9923	0.1199
2018	0.1658	0.9923	0.1199
2019	0.0000	1.0575	0.2264
2020	0.0000	1.0575	0.2264
2021	0.1658	0.9923	0.1199
2022	0.2018	0.9718	0.1036
2023	0.1658	0.9923	0.1199
2024	0.1658	0.9923	0.1199
Average	0.2658	0.9134	0.0854

Further, we checked sensitivity of the results obtained in the WR BoD DEA model to variation in the lower bounds on the weights $v_r \geq w\hat{\beta}_r$, $j = 1, 2, 3$. The results reported in Table 2 are obtained with $w = 0.75$. The relative efficiency scores obtained with $w = 0.10$, 0.50, 0.80 and 0.90 in BoD DEA model are reported in Table 7. The limited sensitivity of the efficiency scores to alternative AR specifications demonstrates that the proposed WR BoD DEA model provides a stable evaluation of the organization's program expense performance. Although tighter weight restrictions reduce discriminatory power of performance,

the resulting changes in efficiency are modest across the weight restrictions and do not materially alter the overall conclusions.

Table 7. Relative efficiency scores obtained under different WR BoD DEA model specifications.

Year (DMU)	Relative Efficiency Scores			
	w = 0.10	w = 0.50	w = 0.80	w = 0.90
2006	0.9621	0.9567	0.9527	0.9514
2007	1	1	1	1
2008	0.9993	0.9972	0.9857	0.9778
2009	0.9802	0.9714	0.9545	0.9446
2010	0.9878	0.9772	0.9588	0.9485
2011	1	1	0.9895	0.9817
2012	1	0.9994	0.9890	0.9830
2013	0.9867	0.9843	0.9734	0.9681
2014	1	1	0.9873	0.9826
2015	0.9681	0.9656	0.9577	0.9543
2016	0.9390	0.9387	0.9330	0.9297
2017	0.9438	0.9437	0.9405	0.9384
2018	0.9656	0.9655	0.9611	0.9581
2019	0.9702	0.9702	0.9652	0.9616
2020	0.9841	0.9835	0.9781	0.9743
2021	0.9435	0.9434	0.9393	0.9367
2022	0.9404	0.9401	0.9358	0.9334
2023	0.9444	0.9443	0.9399	0.9370
2024	0.9481	0.9480	0.9431	0.9398
Average	0.9717	0.9698	0.9624	0.9579

Moreover, when we estimate WR BoD DEA model with w = 0.10, 0.50, 0.80 and 0.90, we find high correlation between the rankings of WR BoD DEA model efficiency scores and the rankings of the BoD DEA model efficiency scores. Overall, the similarity in efficiency trends and rankings suggests that the findings are robust to alternative WR BoD DEA specifications and BoD DEA model. These findings support the robustness of the proposed methodology and suggest that the performance assessment is not driven by a particular choice of weight restrictions.

To examine the extent to which conventional financial performance measures align with efficiency assessments derived from the proposed DEA approach over time, Spearman rank correlations were calculated between the raw PER measures reported in Table 3 and the relative efficiency scores estimated by the WR BoD DEA model reported in Table 4. This comparison is intended to establish the practical contribution of the proposed framework while recognizing the respective strengths and limitations of ratio-based and DEA-based approaches. The results reveal differences in the degree of agreement between the annual rankings obtained from the financial reporting measures and the DEA model. The restated financial data exhibits a positive correlation of 0.56 with the DEA rankings, indicating a moderate degree of consistency in the relative ordering of years. In contrast, the Form 990 rankings demonstrate a weak positive correlation with the DEA rankings (0.37), while the audited rankings exhibit a negative correlation (−0.57). Given that Rank 1 represents the highest efficiency, the negative correlation indicates an inverse association between the annual performance rankings based on audited data and those generated by the WR BoD DEA model. Furthermore, the negative correlation between the audited and restated rankings (−0.68) suggests substantial differences in the relative ordering of years under these two reporting bases. Overall, these findings reveal inconsistencies in the annual performance rankings derived from the conventional PER measures and the WR BoD DEA model. From

a decision-making perspective, the DEA approach provides a systematic, multi-output assessment of relative efficiency across years, complementing conventional PER-based measures and offering additional insights into changes in performance over time.

7. Concluding Remarks, Limitations and Future Research

This study proposes the use of BoD DEA for evaluating three alternative estimates of PER. The motivation is that, when several estimates of PER are available, their relative importance need not be known a priori. The BoD framework provides a natural approach to this problem by allowing the weights assigned to the alternative estimates to be determined endogenously in each year of assessment. However, the flexibility of unrestricted BoD DEA can also create a limitation. Although the three estimates of PER are intended to capture the same underlying construct, the unrestricted model can effectively base the assessment for a particular year on only a subset of the available estimates. The empirical results illustrate this limitation clearly.

To address this issue, this study incorporates weight restrictions into the BoD framework. The restrictions limit the extent to which the relative importance of the alternative estimates can differ while retaining the endogenous weighting principle of BoD DEA. In this way, the proposed approach avoids the complete exclusion of any individual estimate and provides a more balanced aggregation of the available information. The resulting efficiency scores and rankings demonstrate that the choice of weight specification can influence the assessment of relative performance, highlighting the importance of examining weight restrictions rather than relying exclusively on BoD DEA model scores.

The robustness analysis further indicates the extent to which the results depend on the strength of the restrictions imposed. This is important because weight restrictions necessarily involve a degree of judgment concerning the relative importance that should be attached to alternative estimates. Accordingly, the restrictions should not be viewed as producing a universally optimal set of weights, but rather as incorporating the substantive requirement that all estimates should make a meaningful contribution to the assessment.

From an empirical perspective, the proposed BoD DEA framework becomes particularly useful when alternative estimates of PER differ materially and provide potentially conflicting assessments of performance. A single measure may provide an incomplete assessment because it ignores information contained in alternative measures. The BoD approach addresses this issue by combining multiple estimates into a composite performance measure, with the relative importance of each measure determined endogenously within the feasible region defined by the imposed weight restrictions. Therefore, the proposed method provides a systematic approach for evaluating performance when alternative measures of the same construct produce divergent assessments. When alternative PER estimates provide different signals about organizational performance, the WR BoD DEA model provides a mechanism for integrating these differences and identifying the relative position of each year based on the collective information contained in all available PER measures.

The broader contribution of this study is methodological. While the empirical application focuses on alternative estimates of the program expense ratio, the proposed WR BoD DEA framework is applicable to performance evaluation problems in which several measures represent the same underlying construct. Different data sources, accounting standards, joint cost allocation methods and rating agencies may produce different assessments of the same construct. The proposed method ensures that each indicator contributes meaningfully through appropriate weight reductions. Consequently, the proposed methodology provides a more robust and transparent basis for evaluating nonprofit performance.

Limitations and Directions for Future Research

This study has limitations that should be considered when interpreting the findings. First, the analysis is based on a single NPO observed over 19 years, which limits the generalizability of the results to other NPOs. In addition, the WR BoD DEA model provides a relative deterministic assessment of performance. Consequently, the composite scores are dependent on the observed sample and the specified weight restrictions and should be interpreted as relative rather than absolute measures of performance.

Further, the proposed WR BoD DEA model does not explicitly account for changes in organizational structures. Moreover, the analysis does not incorporate external contextual factors that may have influenced the PER estimates over time. As a result, changes in the composite PER scores may not be attributed solely to changes in organizational performance.

Future research could extend the proposed framework to multiple NPOs observed over longer periods, thereby improving generalizability and enabling a more systematic assessment of temporal performance dynamics. Where sufficient comparable panel data are available, window analysis could be used to examine performance across overlapping time periods, while the Malmquist productivity index could be employed to distinguish changes in relative efficiency from shifts in the production frontier.

Incorporating resource-based inputs, funding, and staffing would enable the analysis to distinguish changes in operational scale from changes in performance. Inclusion of contextual variables such as macroeconomic indicators, policy changes, funding levels, and sector-specific developments would provide greater insight into the external factors influencing performance over time. Such extensions would enhance the robustness and explanatory power of WR BoD DEA as a tool for constructing and evaluating composite performance measures.

Author Contributions: Conceptualization, H.H. and D.G.; methodology, D.G.; software, H.H.; validation, D.G. formal analysis, D.G.; investigation, H.H.; resources, H.H.; data curation, H.H.; writing—original draft preparation, H.H. and D.G.; writing—review and editing, H.H.; visualization, D.G.; funding acquisition, H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Social Sciences and Humanities Research Council of Canada, under Insight Grant # number 435-2022-0453.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available to preserve the anonymity of the participating NPO.

Acknowledgments: During the preparation of this manuscript/study, the authors used ChatGPT (OpenAI), for the purpose of language and grammar polishing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Garven, S.A.; Hofmann, M.A.; McSwain, D.N. Playing the numbers game: Program ratio management in nonprofit organizations. *Nonprofit Manag. Leadersh.* **2016**, *26*, 401–416.
2. Finkler, S.A.; Smith, D.L.; Calabrese, T.D.; Purtell, R.M. *Financial Management for Public, Health, and Not-For-Profit Organizations*; CQ Press: Thousand Oaks, CA, USA, 2017.
3. Yetman, M.H.; Yetman, R.J. Do donors discount low-quality accounting information? *Account. Rev.* **2013**, *88*, 1041–1067. [CrossRef]
4. Cashwell, K.; Copley, P.; Dugan, M. Using ratio analysis to manage not-for-profit organizations. *CPA J.* **2019**, *89*, 52–57.
5. Herath, H.S.B.; Herath, T.C. Joint cost allocation anomaly detection in not-for-profit organizations (NPOs) using cluster analysis. *Account. Bus. Public Interest* **2021**, *20*, 444–465.

6. Flack, T.; Hager, M.; Pollak, T.; Wing, K.; Rooney, P. The Pros and Cons of Financial Efficiency Standards. 2004. Available online: https://www.urban.org/sites/default/files/publication/57756/311055-The-Pros-and-Cons-of-Financial-Efficiency-Standards.PDF (accessed on 10 June 2026).
7. Krishnan, R.; Yetman, M.H.; Yetman, R.J. Expense misreporting in nonprofit organizations. *Account. Rev.* **2006**, *81*, 399–420. [CrossRef]
8. Khumawala, S.B.; Parsons, L.M.; Gordon, T.P. Assessing the quality of not-for-profit efficiency ratios: Do donors use joint cost allocation disclosures? *J. Account. Audit. Finance* **2005**, *20*, 287–309. [CrossRef]
9. van der Heijden, H. Charities in competition: Effects of accounting information on donating adjustments. *Behav. Res. Account.* **2013**, *25*, 1–13. [CrossRef]
10. Yetman, R.J. Tax-motivated expense allocations by nonprofit organizations. *Account. Rev.* **2001**, *76*, 297–311. [CrossRef]
11. Schubert, P.; Boenigk, S. Pressure, cost accounting, and resource allocations: Experimental evidence on nonprofit financial decision-making. *Nonprofit Volunt. Sect. Q.* **2021**, *50*, 1216–1238. [CrossRef]
12. Fukuyama, H.; Tan, Y.; Wanke, P. Joint weak disposability in benefit-of-the-doubt model: Methodology and structural analysis of country performance. *Eur. J. Oper. Res.* **2026**, *in press*. [CrossRef]
13. Vakkuri, J. Research techniques and their use in managing non–profit organizations–an illustration of DEA analysis in NPO environments. *Finan. Account. Manag.* **2003**, *19*, 243–263. [CrossRef]
14. Medina-Borja, A.; Triantis, K. Modeling social services performance: A four-stage DEA approach to evaluate fundraising efficiency, capacity building, service quality, and effectiveness in the nonprofit sector. *Ann. Oper. Res.* **2014**, *221*, 285–307. [CrossRef]
15. Miragaia, D.; Brito, M.; Ferreira, J. The role of stakeholders in the efficiency of nonprofit sports clubs. *Nonprofit Manag. Lead.* **2016**, *27*, 113–134. [CrossRef]
16. Quinn, M.; Clerkin, B.; Connolly, C.; Martinez, C. Efficiency of Irish charities: Insights through data envelopment analysis. *Account. Finance Gov. Rev.* **2024**, *33*, 1–29. [CrossRef]
17. Nunamaker, T.R. Using data envelopment analysis to measure the efficiency of non-profit organizations: A critical evaluation. *Manag. Decis. Econ.* **1985**, *6*, 50–58. [CrossRef]
18. Ahn, T.; Charnes, A.; Cooper, W.W. Some statistical and DEA evaluations of relative efficiencies of public and private institutions of higher learning. *Socio-Eco. Plan. Sci.* **1988**, *22*, 259–269. [CrossRef]
19. Medina-Borja, A.; Pasupathy, K.S.; Triantis, K. Large-scale data envelopment analysis (DEA) implementation: A strategic performance management approach. *J. Oper. Res. Soc.* **2007**, *58*, 1084–1098. [CrossRef]
20. Luksetich, W.; Hughes, P.N. Efficiency of fund-raising activities: An application of data envelopment analysis. *Nonprofit Volunt. Sec. Q.* **1997**, *26*, 73–84. [CrossRef]
21. Tofallis, C.; Sargeant, A. Assessing charities using data envelopment analysis. In *Performance Measurement: Past, Present and Future*; Neely, A.D., Ed.; Cranfield University: Bedfordshire, UK, 2000.
22. Berber, P.; Brockett, P.L.; Cooper, W.W.; Golden, L.L.; Parker, B.R. Efficiency in fundraising and distributions to cause-related social profit enterprises. *Socio-Econ. Plan. Sci.* **2011**, *45*, 1–9. [CrossRef]
23. Miragaia, D.A.; Ferreira, J.J.; Vieira, C.T. Efficiency of non-profit organisations: A DEA analysis in support of strategic decision-making. *J. Knowl. Econ.* **2024**, *15*, 3239–3265. [CrossRef]
24. Coupet, J. Exploring the link between government funding and efficiency in nonprofit colleges. *Nonprofit Manag. Leadersh.* **2018**, *29*, 65–81. [CrossRef]
25. Kim, H.; Lee, C.W. Efficiency analysis for nonprofit organizations using DEA: Focused on humanitarian assistance organizations in South Korea. *Asia Pac. J. Innov. Entrep.* **2018**, *12*, 165–180.
26. Coupet, J.; Berrett, J.; Broussard, P.; Johnson, B. Nonprofit benchmarking with data envelopment analysis. *Nonprofit Volunt. Sect. Q.* **2021**, *50*, 647–661. [CrossRef]
27. Cui, C.; Harrison, J.; Ng, F.; Rouse, P. Data envelopment analysis and accounting measures. *Ann. Oper. Res.* **2025**, *351*, 1353–1376. [CrossRef]
28. Garven, S.A.; Beck, A.W.; Parsons, L.M. Are audit related factors associated with financial reporting quality in nonprofit organizations? *Audit.-J. Pract. Theory* **2018**, *37*, 49–68. [CrossRef]
29. Parsons, L.M.; Pryor, C.; Roberts, A.A. Pressure to manage ratios and willingness to do so. *Nonprofit Volunt. Sect. Q.* **2017**, *46*, 1–20. [CrossRef]
30. Coupet, J.; Berrett, J.L. Toward a valid approach to nonprofit efficiency measurement. *Nonprofit Manag. Leadersh.* **2019**, *29*, 299–320. [CrossRef]
31. Pina, V.; Torres, L.; Bachiller, P. Political influence and the performance of nonprofit Spanish banks. *Nonprofit Manag. Leadersh.* **2016**, *26*, 471–488. [CrossRef]
32. Thanassoulis, E.; Boussofiane, A.; Dyson, R.G. A comparison of data envelopment analysis and ratio analysis as tools for performance assessment. *Omega* **1996**, *24*, 229–244. [CrossRef]
33. Feroz, E.H.; Kim, S.; Raab, R.L. DEA as a complement to financial ratio analysis. *J. Oper. Res. Soc.* **2003**, *54*, 48–58.

34. Chen, W.C.; McGinnis, L.F. Reconciling ratio analysis and DEA as performance assessment tools. *Eur. J. Oper. Res.* **2007**, *178*, 277–291. [CrossRef]
35. Almulhima, T.; Almubarak, N.; Aljabr, N. How to comprehensively evaluate firm performance from operational, financial, and sustainability perspectives? A two-stage data envelopment analysis approach. *Emerg. Mark. Finance Trade* **2024**, *60*, 1447–1467. [CrossRef]
36. Keating, E.K.; Frumkin, P. Reengineering nonprofit financial accountability: Toward a more reliable foundation for regulation. *Public Adm. Rev.* **2003**, *63*, 3–15. [CrossRef]
37. Charnes, A.; Cooper, W.W.; Rhodes, E. Measuring the efficiency of decision-making units. *Eur. J. Oper. Res.* **1978**, *2*, 429–444. [CrossRef]
38. Cooper, W.W.; Seiford, L.M.; Zhu, J. (Eds.) *Handbook on Data Envelopment Analysis*, 2nd ed.; Springer: New York, NY, USA, 2011.
39. Banker, R.D.; Charnes, A.; Cooper, W.W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manag. Sci.* **1984**, *30*, 1078–1092. [CrossRef]
40. Cherchye, L.; Moesen, W.; Rogge, N.; Van Puyenbroeck, T. An introduction to 'Benefit of the Doubt' composite indicators. *Soc. Indic. Res.* **2007**, *82*, 111–145. [CrossRef]
41. Sing, R.K.; Sharma, S.K. A non-parametric synthesize measure for corporate governance: Empirical evidence from Indian banks. *Int. Rev. Econ. Finance* **2023**, *88*, 258–279. [CrossRef]
42. Lovell, C.A.K.; Pastor, J.T.; Turner, J.A. Measuring macroeconomic performance in the OECD: A comparison of European and non-European countries. *Eur. J. Oper. Res.* **1995**, *87*, 507–518. [CrossRef]
43. Rogge, N. On aggregating benefit of the doubt composite indicators. *Eur. J. Oper. Res.* **2018**, *264*, 364–369. [CrossRef]
44. Hasannasab, M.; Margaritis, D.; Rouse, P. Benefit of the doubt models: Some issues and insights. *J. Product. Anal.* **2026**, *65*, 14. [CrossRef]
45. Fugger, G. Absolute and relative weight restrictions in DEA-a comparison. *Int. J. Oper. Res.* **2023**, *47*, 245–271. [CrossRef]
46. Thompson, R.G.; Singleton, F.D.; Thrall, R.M.; Smith, B.A. Comparative site evaluations for locating a high-energy physics lab in Texas. *Interfaces* **1986**, *16*, 35–49. [CrossRef]
47. Wong, Y.-H.B.; Beasley, J.E. Restricting weight flexibility in DEA. *J. Oper. Res. Soc.* **1990**, *41*, 829–835. [CrossRef]
48. Dyson, R.G.; Thanassoulis, E. Reducing weight flexibility in data envelopment analysis. *J. Oper. Res. Soc.* **1988**, *39*, 563–576. [CrossRef]
49. Puyenbroeck, T.V. On the output orientation of the benefit-of-the-doubt-model. *Soc. Indic. Res.* **2018**, *139*, 415–431. [CrossRef]
50. Camanho, A.S.; Zanella, A.; Moutinho, V. Benefit-of-the-doubt composite indicators and use of weight restrictions. In *Advanced Mathematical Methods for Economic Efficiency Analysis*; Springer: Cham, Switzerland, 2023; pp. 93–112.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.