Nonprofit and Voluntary Sector Quarterly
1–15

DOI: 10.1177/0899764020977670
journals.sagepub.com/home/nvs
SAGE

Nonprofit Benchmarking With Data Envelopment Analysis

Jason Coupet[1], Jessica Berrett[2], Paul Broussard[1], and Bradley Johnson[1]

Abstract
Benchmarking nonprofit performance can be challenging, constraining the ways nonprofits can use operational data to learn from each other and highlight organizational progress. Valid output or outcome data are scarce, and there is much to learn about measuring performance even when these data are available. Data envelopment analysis (DEA) is a mathematical linear programming technique that can be used to measure performance in a way that not only produces valid efficiency scores but also allows for benchmarking nonprofits with similar service missions. Using financial and production data from the nonprofit transportation sector, we walk through an example to explore DEA as a tool to measure and benchmark nonprofits. We conclude with suggestions for practice, emphasizing that DEA is useful for stakeholders looking to benchmark nonprofits by underscoring production and performance.

Keywords
benchmarking, productivity, data envelopment analysis, strategic management, performance

Pressing questions loom about nonprofit performance management theory (Mitchell & Calabrese, 2020), and how stakeholders benchmark nonprofits is a part of the conversation (Mitchell & Calabrese, 2019). Benchmarking, or the process of identifying top performers and adopting the best practices from identified strategic leaders, can be key

[1]North Carolina State University, Raleigh, USA
[2]University of Colorado–Colorado Springs, USA

Corresponding Author:
Jason Coupet, Associate Professor of Public Administration, School of Public and International Affairs, North Carolina State University, 2221 Hillsborough St., 229 Caldwell Hall, Raleigh, NC 27695, USA.
Email: jacoupet@ncsu.edu

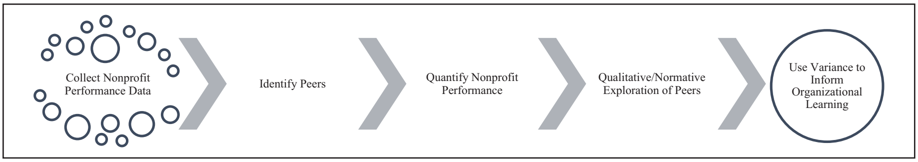

Figure 1. Benchmarking for learning.

to organizational learning (Ammons, 2014; Askim et al., 2007). The most useful benchmarking practices involve "rigorous procedures for identifying best-in-class performers, analyzing processes, and isolating elements that contribute to excellent results" (Ammons & Roenigk, 2014, p. 314).

The most popular nonprofit benchmarking approach by third parties involves comparing nonprofits with financial ratios (Hager et al., 2001; Prentice, 2016). Financial ratios somewhat capture spending patterns, but they do not capture what nonprofits *do* with resources, leaving out outputs and outcomes entirely. Only focusing on inputs can encourage nonprofits to forgo necessary investments to appear leaner to donors (Krishnan et al., 2006; Parsons et al., 2017).

There are recent proposals for holistic nonprofit benchmarking approaches, with several scholars proposing multidimensional measures of nonprofit performance including balanced scorecards (Kaplan, 2001) and other approaches (Liket & Maas, 2015; Sowa et al., 2004) including nonprofit performance assessment tools that are designed to engage nonprofits in best practices from nonprofit managers, boards, and partnerships (Herman & Renz, 1999).

Robust nonprofit benchmarking techniques should consider not only the resources nonprofits use but also what they *produce*. They should be continuous and processes oriented and place emphasis on learning (see Figure 1) from other organizations (Ammons, 2014). DEA uses inputs and outputs to identify top performers and group peers and has widely been used as a benchmarking tool outside of nonprofit management (Donthu et al., 2005; Emrouznejad & Yang, 2018). However, its use in nonprofit management has been limited.

DEA

Data envelopment analysis (DEA) has been used by some scholars as a benchmarking tool in a dual pronged approach where peers identified and scored with DEA and are investigated with balanced scorecards or another approach in the second stage designed to investigate best practices (Shafiee et al., 2014). DEA uses a nonparametric, deterministic linear programming methodology that estimates a single, synthesized measure of an organization's performance. The singular performance measure is estimated by maximizing the sum of each organization's output to input ratio, denoted by θ in Figure 2. DEA has its roots in organizational economics and was initially developed to measure efficiency in public service organizations where singular measures of outputs or outcomes might prove too simple.

$$\max_{u,v} \quad \theta = \frac{\sum_{m=1}^{M} u_m y_{mk'}}{\sum_{n=1}^{N} v_n x_{nk'}} \tag{2.1}$$

$$\text{subject to} \quad 1 \geq \frac{\sum_{m=1}^{M} u_m y_{mj}}{\sum_{n=1}^{N} v_n x_{nj}} \quad \forall j \tag{2.2}$$

$$\sum_{n=1}^{N} v_n x_{nk'} = 1 \tag{2.3}$$

$$u_m, v_n, y_{mj}, x_{nj} > 0 \quad \forall m, n, j \tag{2.4}$$

Figure 2. Data envelopment analysis.

In short, DEA uses weighted ratios of inputs (what nonprofits use to produce or serve) and outputs (a quantified measure of what nonprofits produce or serve) to compose a performance measure for each member of a set of nonprofits. This score is used to benchmark nonprofits and identify peers from which each nonprofit can learn.

In Figure 2, *m* represents the firm or institutional unit; *y* and *x* represent outputs and inputs, respectively; and *u* and *v* represent output and input weights, respectively. In Equation 2.1 (see Figure 2), DEA computes a performance score (θ) using weighted inputs and outputs for decision making unit (DMU) *k′*, a target nonprofit. It is a maximization program that finds the combination of weights for each input and output that creates the highest performance score θ. For example, if a transportation nonprofit does not make many trips but drives many miles, DEA will heavily weight the ratios that include the number of miles driven.[1]

The top performing organizations receive a score of 1 (usually represented as a percentage), and this score "envelops" that of lower performing peers for useful comparisons. DEA envelops nonprofit performance by comparing each nonprofit to its better performing peers, so each performance score is computed relative to others' performances. DEA defines peers as collectives whose performance scores are linked. In short, nonprofits are peers if they are enveloped by the same top performer. DEA restricts this performance score (θ) to a value between 0 and 1 (unity) for all nonprofits j. This is evident in Equations 2.3 and 2.4 (see Figure 2), respectively, where we set the sum of all weighted inputs to 1 and set all inputs and outputs to positive numbers. Figure 2 shows that a nonprofit's performance is determined by the weighted sum of its outputs (what it might produce) divided by the weighted sum of its inputs (what it might use toward its missions). DEA then selects the weights that maximize each nonprofit's score. This performance score can be used by scholars as a dependent variable to measure the impact of exogenous variables (i.e., building theory to predict nonprofit performance). DEA has been successfully deployed in the management and public management discipline. It has had less use in nonprofit management, but there have been some examples of scholarship where DEA is operationalized as a key performance dependent variable in nonprofit research. This is key since researchers have sometimes lacked valid output-based measure of performance (Bryan, 2019; Coupet, 2018; Ford & Ihrke, 2020; Fulton, 2020; Luksetich & Hughes, 1997; Miragaia et al., 2016).

The top performing organizations form a frontier based on their input/output combinations, as DEA computes performance scores *and* envelops those scores with the scores of high-performing peers, constructing convenient targets for more qualitative or contextual best practices for organizational learning. As Cook et al. (2014, p. 1) note, "it must be remembered that ultimately DEA is a method for performance evaluation and benchmarking against best-practice."

Much of DEA's appeal as a benchmarking tool is a function of its process for identifying peers. DEA groups nonprofits by using the data to determine which ones have "the closest similar circumstances as defined by the set of inputs and outputs" (Jain et al., 2011). Figure 3 graphically illustrates how DEA finds peers with hypothetical Nonprofits A, B, C, and D. Nonprofits A, B, and C lie on the production possibility frontier (PPF)[2] as the top performing nonprofits. If Nonprofit D were to improve performance with the same mix of resources and output, it would reach the frontier at D′.[3] D′ exists only potentially, so Nonprofit D's hypothetical path to improvement lies on a frontier point between Nonprofits A and B. Any nonprofit in the triangle OAB would have Nonprofits A and B as peer benchmarks.

Executing DEA

There are several steps to executing DEA. First, a sample must be carefully selected. Nonprofits should have similar missions and very similar production functions. Second, inputs and outputs should be selected. Inputs and outputs should be vetted by rigorous management processes and be agreed upon by all stakeholders that plan to strategically use DEA. Outputs should be quantifiable and strategically important, while inputs, also quantifiable, should theoretically increase outputs or improve outcomes as they increase (Coupet, 2018). For instance, an increase in the number of volunteer drivers, holding other factors constant, should increase the

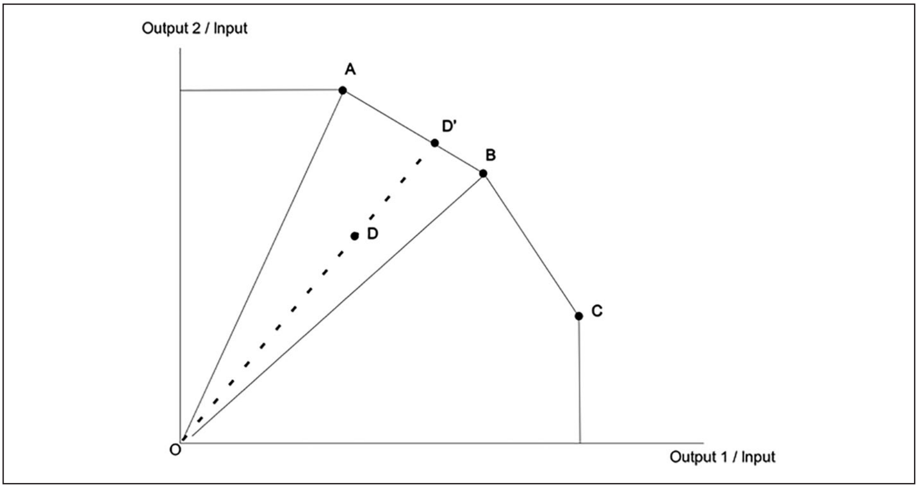

Figure 3. DEA peers.

number of trips each nonprofit can make. Third, the input and output data should be adjusted for missing values, measurement error, and outliers. Since DEA is an envelopment methodology, unreliable data can affect other scores if used to build a frontier.

Fourth, since size can provide scale advantages (or disadvantages), and most DEA programs can adjust scores to account for scale effects, researchers using DEA should test for scale effects by regressing outputs on squared input terms (Banker et al., 1988). They might also consider restricting the DEA weights if a particular mission or resource is critically important. There are several ways to execute the DEA program; well-known statistical software, like STATA (Ji & Lee, 2010) and R (Wilson, 2008), have DEA procedures and readily available packages.[4] Table 1 provides succinct guidance for researchers using DEA.

Example: Nonprofit Rural Transit

Data

To illustrate DEA as a benchmarking tool for nonprofits, we explore a sector both with quantifiable inputs and outputs and a consensus about the meaningful inputs and outputs (Daraio et al., 2016). We examined 29 U.S. rural transportation nonprofits, which are largely rural buses or shuttles administered to transport disadvantaged rural populations where public transportation is scarce or nonexistent.

While several more nonprofits report data to the Federal Transportation Administration (FTA), only 29 had complete volunteer and financial data for 2015. We deidentify the data here although it is public because this example is for illustration purposes only. This is important given our sample, and we wish only to illustrate how

Table 1. Executing Data Envelopment Analysis.

Step	Action	Considerations
1	Carefully select a sample of nonprofits	Nonprofits should have very similar missions and production functions
2	Carefully select inputs and outputs	Inputs and outputs should be relevant to the nonprofit mission, and all inputs should, on average, increase all outputs. This can be tested with regression techniques
3	Adjust for missing data/ measurement error	DEA scores are particularly sensitive to outliers and data errors
4	Assess returns to scale	Most DEA programs allow researchers to adjust for increasing returns to scale
5	Consider weight restrictions	Weights can be adjusted in most DEA programs if a particular mission or resource is critical
6	Execute the DEA linear program	Easy-to-use packages developed in most statistical software like STATA, R, and Excel
7	Investigate peer groups for organizational learning	Work with stakeholders to use DEA benchmarks to facilitate further organizational learning

the data might be used to benchmark nonprofits. We do not attempt to draw any theoretical conclusions about nonprofit performance in the rural transportation space.

We take advantage of a unique data reporting structure in this study. Many rural transit operations are often contracted to nonprofit organizations when areas are hard to reach for large public metropolitan transportation organizations. As subcontractors, nonprofits report a subset of operational and financial data to the FTA, meeting some of the same reporting requirements as larger public transit agencies. The nonprofits that do not directly report to public sector organizations receive Section 5311 funding from the FTA, which helps non-urban transit nonprofits with operational costs (Koffman et al., 2004).

Inputs and Outputs

Input and output selection is critical in DEA and should be done carefully (Dyson et al., 2001). Inputs should be valid and reliable measurements of the costs and resources nonprofits use to achieve their missions, and outputs should be valid and reliable measurements of what the nonprofit produces with those resources. Researchers should rely either on a deep body of literature validating inputs and outputs or on statistically validating the nonprofit production function, or they should work closely with the set of nonprofits being benchmarked.[5]

We chose this subset of nonprofits because the transportation management literature has achieved relative clarity among inputs and outputs. Our inputs include vehicles in annual maximum service, number of volunteer drivers, and operational expenses. Outputs include car miles and unique passenger trips. The DEA benchmarking literature in transportation is vast and includes thousands of studies,[6] and our inputs and outputs are among the most commonly used in the transportation literature wherein DEA is used as a benchmarking tool (Daraio et al., 2016).

Results and Interpretation

We use EMS to run our DEA linear program[7] and compute performance scores for 2015. In Table 2, we walk through a snapshot of peers as an illustration. DEA computed that Nonprofit 3 was on the performance frontier and Nonprofits 23 and 28 were peers. Note that Nonprofits 3, 23, and 28 have similar numbers of vehicles and volunteers. Nonprofit 3 lies on the performance frontier and in 2015 had a fleet of 6 vehicles, 14 volunteers, and about $800,000 in expenses. With these resources, it drove more than 400,000 miles in just over 20,000 trips. Nonprofit 23 was about 40% (its performance score) as productive as Nonprofit 3 (its peer) because of having more fleet vehicles (7) and volunteers (17) but drove approximately one fifth of the miles and half the number of trips. Ideally, any factors or nuance driving the performance difference could be parsed out in a later qualitative or normative exploration and best practices could be adapted.

Nonprofits 23 and 28 show the inherent problems involved with overemphasizing expenses. With unobservable output, Nonprofit 3 has much higher expenses and might appear quite unattractive to donors. Nonprofits 23 and 28 have used similar numbers of vehicles and volunteers as Nonprofit 3, but Nonprofit 3 is on the performance frontier because it has driven nearly twice the number of trips as Nonprofit 23 and nearly 10 times the miles as Nonprofit 28. Nonprofits 23 and 28 drive many fewer miles per trip than Nonprofit 3, so one might infer that lower gasoline costs (likely to rise with miles and not trips, ceteris paribus) are causing the lower expenses for Nonprofits 23 and 28, not necessarily more efficient management. Stakeholders or donors looking only at expenses, volunteers, and vehicles might penalize Nonprofit 3, while DEA finds (using inputs *and* measures of the services that Nonprofit 3 produces) that Nonprofit 3 is a peer from which others might learn.

DEA Benchmarking Strengths

DEA has several strengths as a benchmarking technology. First, it typically works well to benchmark nonprofits that use many kinds of resources to achieve many different goals. The linear program in Figure 2 can sum up as many input/output ratios as an evaluator's computing power will allow and uses this weighted sum to compute a single performance score for each of a set of nonprofits that can be used to compare general performance. Nonprofit organizations often manage multiple goals (Herman & Renz, 1999), and DEA can help address the multiple mission dilemma in nonprofit benchmarking. For instance, consider that nonprofits in our sample make short and long trips, with DEA accounting for both in a single performance measure that can be unpacked for organizational learning.

Table 2. Example—Results Snapshot.

							Weights				
							Inputs			Outputs	
Nonprofit	Score	Vehicles	Volunteers	Expenses	Miles	Trips	Vehicles	Volunteers	Expenses	Miles	Trips
3	100.00%	6	14	$809,514.00	408,558.00	20,163	1	0	0	1	0
23	40.10%	7	17	$238,997.00	82,236.00	10,696	0.73	0	0.27	0.63	0.37
28	56.82%	5	12	$201,032.00	46,699.00	16,621	1	0	0	0.12	0.88

Second, DEA uses variable weights not only to compose performance scores but to also assign peers. Since the weights reflect the resources each nonprofit uses best and the outputs or outcomes it emphasizes most, DEA envelops performance scores with nonprofits with similar input and output mixes. Nonprofits, for instance, that use more volunteers than staff to package meals are likely to be grouped as peers. Most DEA software packages explicitly note which organizations are peers of the highest performing nonprofits. DEA provides a scientific way to identify peers from which organizational learning can occur, either by examining the data and results from DEA or by using the analysis to identify other nonprofits for more nuanced learning. The weights selected by DEA can also inform nonprofits with multiple missions about what they do best. Highly weighted inputs and outputs can indicate that a nonprofit uses a particular resource or achieves a particular mission well.

Moreover, researchers do not need to identify any particular functional form for benchmarking (Charnes et al., 1981). This is important for nonprofits juggling multiple goals, where it can be difficult to decipher what resources used are designed for which particular outcomes and outputs nonprofits take on. Rather, DEA computes a performance score by allowing weights to vary to maximize performance. Resources used and outcomes achieved well will be heavily weighted for the nonprofits being benchmarked.

Limitations of DEA

Scholars should exercise several points of caution when using DEA as a strategic planning tool. First, DEA-based benchmarks are only as strategically useful as the inputs and outputs/outcomes selected. Recall that DEA weights inputs and outputs to compose performance scores, so haphazardly selecting inputs and outputs can artificially inflate or deflate nonprofit performance and can yield results generally far less useful for strategic purposes.

Second, DEA does not provide qualitative information about performance. Peers might have environmental conditions not easily quantifiable. For instance, an efficient peer might have less cumbersome permit processes or strategic partnerships that allow for more productive operations, and less efficient peers might have more expensive (or rare) inputs for reasons beyond the unit's control. Thus, DEA is best used in conjunction with other strategic planning and benchmarking techniques like interviews and discussion with managers and peers. Since nonprofits often have limited capacity for performance management (Lee & Clerkin, 2017), we acknowledge the DEA might ultimately be most useful for self-study and for others with more intimate sector knowledge.

Third, DEA is *not* statistical. It does not model error terms, so researchers cannot rely on the probabilistic principles of statistical inference to strengthen validity. Performance scores generated by DEA are therefore more concrete, highlighting the extreme caution with which inputs and outputs should be selected. Researchers should also carefully examine data for measurement errors since erroneous outliers might envelop and affect other scores. DEA processes can (and often should) be augmented by statistical procedures, but researchers using DEA should be comfortable that DEA scores, weights, or benchmarks are not themselves estimated parameters. DEA scores do, however, typically perform well as dependent or independent variables (McDonald, 2009).

Last, DEA scores can be highly sensitive to outliers since they are developed by measuring the distance from each nonprofit to its most efficient peers. One extremely high-performing nonprofit can make other nonprofits appear to perform worse than they actually do, again highlighting the extreme care that should be taken in selecting inputs and outputs and ensuring each nonprofit in the benchmarking sample is highly similar both in their missions and the ways in which they achieve it. For instance, if one out of several nonprofits in a sample does not use volunteers as a matter of principle, it can appear artificially highly efficient since the input value of 0 will be weighted heavily. A similar logic applies to sample sizes; a small sample size relative to the number of inputs and outputs means that a high enough number of organizations might appear on the frontier that peer comparisons are no longer useful. A rule of thumb for DEA is that the sample size should equal the number of inputs times the number of outputs times three.[8]

Implications

The DEA technique has some advantages for external stakeholders looking to identify high-performing charities. DEA overcomes some of the well-known problems of financial ratios and is more useful than reputational assessments for cross-organizational comparisons in the same sector. It uses *both* the resources charities use *and* what charities do with those resources to benchmark performance and identify peers, enabling nonprofits to move beyond identifying peer benchmarks with reputation or financial ratios alone, potentially avoiding the starvation cycle by emphasizing production (Lecy & Searing, 2015).

DEA has had limited use as a nonprofit benchmarking tool, and scarcity of data might play a role in the past. As Callen et al. (2003) note,

> Unfortunately, more theoretically defensible production theoretic measures of efficiency, such as those derived from stochastic frontier estimation procedures or data envelopment analysis, require input-output data that are simply unavailable for all but a very few U.S. nonprofits. (p. 517)

In the past two decades, however, the availability and emphasis on data collection and outcome management has grown (Benjamin, 2012; Mitchell & Berlan, 2016) and interest in building output and outcome data has increased in government organizations that support the sector (Fitzgerald et al., 2019; Urban Institute, 2006, 2020). DEA could supplement these efforts in instances where stakeholders can use *valid* resource and output measures by which nonprofits can credibly be evaluated, such as with charities that contract with governments to produce quantities of goods and services that can be credibly counted like home building or meal packaging.

DEA does not directly indicate whether the differences between peer nonprofits are due to managerial factors or environmental conditions beyond the scope of managerial control. Environmental conditions and managerial best practices are best explored with further inquiry, which is key to any benchmarking practice geared toward organizational learning. Applications of DEA that involve environmental variables, as well as other benchmarking techniques including environmental variation, are referenced in Coelli (1996) and Coupet (2018).

There have been several recent scholarly efforts to improve indicators of nonprofit performance (Coupet & Berrett, 2019; Donthu et al., 2005; Mitchell, 2017). We see this study as part of a larger initiative to better map management science to the norms of nonprofit strategic management practices (Mitchell & Calabrese, 2019, 2020). Identifying best practices should be based on valid empirical indications that emphasize performance indicators as more useful than financial ratio benchmarking and indicators of fiscal leanness. DEA is a benchmarking tool that uses nonprofit inputs but also emphasizes what nonprofits *do* with these resources. As third-party organizations and other nonprofit data platforms look beyond financial ratios for valid nonprofit benchmarking tools, techniques like DEA might help others learn from nonprofits and help nonprofits learn from each other and themselves by emphasizing how DEA can help future research better map management science to the norms of nonprofit strategic management practices. We argue that DEA possesses untapped potential as an evaluation tool, allowing for organizations to compare themselves with other, similar organizations. It compliments other benchmarking tools, and most importantly, can help nonprofits and stakeholders move beyond learning from peers using reputation and financial ratios. DEA can support the movement toward more outcome- and output-oriented management practices and a management culture that is more focused on outcomes and outputs, not just costs.

Acknowledgments

We would like to thank three gracious reviewers, the editorial team, and ARNOVA scholars for contributions that substantially improved this manuscript. We would also like to thank Rich Clerkin, the Institute for Nonprofits at NC State, and the Research Office in the College of Humanities and Social Sciences at NC State for resources that assisted in this study.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Jason Coupet https://orcid.org/0000-0002-7006-3598

Notes

1. Data envelopment analysis (DEA) weights can be programmed as fixed if inputs and outputs are constrained in some meaningful way (Barnum et al., 2017). Inputs and outputs should be positive, and negative outcomes should be transformed. There is a robust literature on DEA with undesirable outputs and inputs (see Seiford and Zhu, 2002).
2. Production possibility frontiers (PPFs) typically represent the maximum possible production given a set on inputs. DEA constructs the PPF with top performers.
3. The distance from D to D′ is known as a "radial distance."
4. Researchers skilled in programming languages can program Microsoft Excel solvers to execute DEA. Efficiency measurement system (EMS), produced by Tim Coelli, is a commonly used free package. There are also several DEA solver packages that evaluators and scholars can buy. In most packages, researchers can simply upload a spreadsheet (or similar data file) with the input and output values for each nonprofit they evaluate. Then, the linear program in Figure 2 (or some close user-specified variant) is executed, resulting in performance scores and identification of similar peers.
5. This can be done by an appropriately modeled regression confirming that an increase in each input, on average, leads to an increase of each output.
6. A Google Scholar search conducted on December 30, 2018, using keywords such as "data envelopment analysis transit" produced 52,000-plus results.
7. Nonprofits performance may depend on economies of scale: their size or the goods and services volume they produce might affect their productivity. Sometimes large organizations can perform better than peers because they can take advantage of their size; other times being too large can hold them back. We assume constant returns to scale in our analysis for simplicity, but this can be tested. See Barnum et al. (2017) for a discussion.
8. DEA most heavily weights the input/output ratios that nonprofits produce the highest nonprofit performance scores. A small sample size relative to the number of possible weighted ratios can mean almost all of the nonprofits appear on the frontier.

References

Ammons, D. (2014). *Municipal benchmarks: Assessing local performance and establishing community standards*. Routledge.

Ammons, D. N., & Roenigk, D. J. (2014). Benchmarking and interorganizational learning in local government. *Journal of Public Administration Research and Theory*, *25*(1), 309–335.

Askim, J., Johnsen, Å., & Christophersen, K. A. (2007). Factors behind organizational learning from benchmarking: Experiences from Norwegian municipal benchmarking networks. *Journal of Public Administration Research and Theory*, *18*(2), 297–320.

Banker, R. D., Charnes, A., Cooper, W. W., & Maindiratta, A. (1988). A comparison of DEA and translog estimates of production frontiers using simulated observations from a known technology. In A. Dogramaci & R. Fare (Eds.), *Applications of modern production theory: Efficiency and productivity* (pp. 33–55). Springer.

Barnum, D., Coupet, J., Gleason, J., McWilliams, A., & Parhankangas, A. (2017). Impact of input substitution and output transformation on data envelopment analysis decisions. *Applied Economics*, *49*(15), 1543–1556.

Benjamin, L. M. (2012). Nonprofit organizations and outcome measurement: From tracking program activities to focusing on frontline work. *American Journal of Evaluation*, *33*(3), 431–447.

Bryan, T. K. (2019). Toward a contingency model for the relationship between capacity and effectiveness in nonprofit organizations. *Nonprofit and Voluntary Sector Quarterly*, *48*(4), 885–897.

Callen, J. L., Klein, A., & Tinkelman, D. (2003). Board composition, committees, and organizational efficiency: The case of nonprofits. *Nonprofit and Voluntary Sector Quarterly*, *32*(4), 493–520.

Charnes, A., Cooper, W. W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. *Management Science*, *27*(6), 668–697.

Coelli, T. (1996). A guide to DEAP version 2.1: A data envelopment analysis (computer) program. *Centre for Efficiency and Productivity Analysis, University of New England*, *96*(08).

Cook, W. D., Tone, K., & Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *Omega*, *44*, 1–4.

Coupet, J. (2018). Exploring the link between government funding and efficiency in nonprofit colleges. *Nonprofit Management and Leadership*, *29*(1), 65–81.

Coupet, J., & Berrett, J. L. (2019). Toward a valid approach to nonprofit efficiency measurement. *Nonprofit Management and Leadership*, *29*(3), 299–320.

Daraio, C., Diana, M., Di Costa, F., Leporelli, C., Matteucci, G., & Nastasi, A. (2016). Efficiency and effectiveness in the urban public transport sector: A critical review with directions for future research. *European Journal of Operational Research*, *248*(1), 1–20.

Donthu, N., Hershberger, E. K., & Osmonbekov, T. (2005). Benchmarking marketing productivity using data envelopment analysis. *Journal of Business Research*, *58*(11), 1474–1482.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, *132*(2), 245–259.

Emrouznejad, A., & Yang, G. L. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-economic Planning Sciences*, *61*, 4–8.

FitzGerald, C., Carter, E., Dixon, R., & Airoldi, M. (2019). Walking the contractual tightrope: A transaction cost economics perspective on social impact bonds. *Public Money & Management*, *39*(7), 458–467.

Ford, M. R., & Ihrke, D. M. (2020). Connecting group dynamics, governance, and performance: Evidence from charter school boards. *Nonprofit and Voluntary Sector Quarterly*, *49*, 1035–1057.

Fulton, B. R. (2020). Bridging and bonding: Disentangling two mechanisms underlying the diversity–performance relationship. *Nonprofit and Voluntary Sector Quarterly*. Advance online publication. https://doi.org/10.1177/0899764020939654

Hager, M. A., Pollak, T., & Rooney, P. (2001). Variations in overhead and fundraising efficiency measures: The influence of size. *Age, and Subsector*. https://scholarworks.iupui.edu/bitstream/handle/1805/6389/variations_in_overhead_and_fundraising_efficiency_measures.pdf?sequence=1&isAllowed=y

Herman, R. D., & Renz, D. O. (1999). Theses on nonprofit organizational effectiveness. *Nonprofit and Voluntary Sector Quarterly*, *28*(2), 107–126.

Jain, S., Triantis, K. P., & Liu, S. (2011). Manufacturing performance measurement and target setting: A data envelopment analysis approach. *European Journal of Operational Research*, *214*(3), 616–626.

Ji, Y. B., & Lee, C. (2010). Data envelopment analysis. *The Stata Journal*, *10*(2), 267–280.

Kaplan, R. S. (2001). Strategic performance measurement and management in nonprofit organizations. *Nonprofit Management and Leadership*, *11*(3), 353–370.

Koffman, D., Raphael, D., & Weiner, R. (2004). *The impact of federal programs on transportation for older adults* (Vol. 17). AARP Public Policy Institute.

Krishnan, R., Yetman, M. H., & Yetman, R. J. (2006). Expense misreporting in nonprofit organizations. *The Accounting Review*, *81*(2), 399–420.

Lecy, J. D., & Searing, E. A. (2015). Anatomy of the nonprofit starvation cycle: An analysis of falling overhead ratios in the nonprofit sector. *Nonprofit and Voluntary Sector Quarterly*, *44*(3), 539–563.

Lee, C., & Clerkin, R. M. (2017). The adoption of outcome measurement in human service nonprofits. *Journal of Public and Nonprofit Affairs*, *3*(2), 111–134.

Liket, K. C., & Maas, K. (2015). Nonprofit organizational effectiveness: Analysis of best practices. *Nonprofit and Voluntary Sector Quarterly*, *44*(2), 268–296.

Luksetich, W., & Hughes, P. N. (1997). Efficiency of fund-raising activities: An application of data envelopment analysis. *Nonprofit and Voluntary Sector Quarterly*, *26*(1), 73–84.

McDonald, J. (2009). Using least squares and tobit in second stage DEA efficiency analyses. *European Journal of Operational Research*, *197*(2), 792–798.

Miragaia, D., Brito, M., & Ferreira, J. (2016). The role of stakeholders in the efficiency of nonprofit sports clubs. *Nonprofit Management and Leadership*, *27*(1), 113–134.

Mitchell, G. E. (2017). Fiscal leanness and fiscal responsiveness: Exploring the normative limits of strategic nonprofit financial management. *Administration & Society*, *49*(9), 1272–1296.

Mitchell, G. E., & Berlan, D. (2016). Evaluation and evaluative rigor in the nonprofit sector. *Nonprofit Management and Leadership*, *27*(2), 237–250.

Mitchell, G. E., & Calabrese, T. D. (2019). Proverbs of nonprofit financial management. *American Review of Public Administration*, *49*(6), 649–661.

Mitchell, G. E., & Calabrese, T. D. (2020). *Instrumental philanthropy, nonprofit theory, and information costs*. Nonprofit Policy Forum, 11, 15.

Parsons, L. M., Pryor, C., & Roberts, A. A. (2017). Pressure to manage ratios and willingness to do so: Evidence from nonprofit managers. *Nonprofit and Voluntary Sector Quarterly*, *46*(4), 705–724.

Prentice, C. R. (2016). Why so many measures of nonprofit financial performance? Analyzing and improving the use of financial measures in nonprofit research. *Nonprofit and Voluntary Sector Quarterly*, *45*(4), 715–740.

Seiford, L. M., & Zhu, J. (2002). Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research*, *142*(1), 16–20.

Shafiee, M., Lotfi, F. H., & Saleh, H. (2014). Supply chain performance evaluation with data envelopment analysis and balanced scorecard approach. *Applied Mathematical Modelling*, *38*(21–22), 5092–5112.

Sowa, J. E., Selden, S. C., & Sandfort, J. R. (2004). No longer unmeasurable? A multidimensional integrated model of nonprofit organizational effectiveness. *Nonprofit and Voluntary Sector Quarterly*, *33*(4), 711–728.

Urban Institute. (2006). https://www.urban.org/sites/default/files/publication/43036/411404-Building-a-Common-Outcome-Framework-To-Measure-Nonprofit-Performance.PDF

Urban Institute. (2020). https://www.urban.org/policy-centers/cross-center-initiatives/performance-management-measurement/projects/nonprofit-organizations/projects-focused-nonprofit-organizations/outcome-indicators-project

Wilson, P. W. (2008). FEAR: A software package for frontier efficiency analysis with R. *Socio-Economic Planning Sciences*, *42*(4), 247–254.

Author Biographies

Jason Coupet is an associate professor of Public Administration in the School of Public and International Affairs at North Carolina State University. His PhD is in Strategic Management from the University of Illinois at Chicago, and his BA in Economics from the University of Michigan. His research interests include strategic management, data envelopment analysis, organizational economics, research methods, and the political economy of organizations.

Jessica Berrett is an assistant professor in the School of Public Affairs at the University of Colorado–Colorado Springs. Her research focuses on nonprofit management including financial management and nonprofit effectiveness. She is currently working on research related to overhead ratios, efficiency, and revenue diversification.

Paul Broussard is a doctoral candidate at NCSU, with areas of interest in performance management, economic development, and public service motivation.

Bradley Johnson is a doctoral candidate in Public Administration and NC State. His research interests are in public involvement implementation, organizational governance, technology, and the intersection of all three.