

FURTHER IMPROVEMENTS OF FINITE SAMPLE APPROXIMATION OF CENTRAL LIMIT THEOREMS FOR ENVELOPMENT ESTIMATORS

Léopold Simar, Valentin Zelenyuk, Shirong Zhao

REPRINT | 2023 / 08

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

<https://uclouvain.be/en/research-institutes/lidam/isba/publication.html>

Further Improvements of Finite Sample Approximation of Central Limit Theorems for Envelopment Estimators

LÉOPOLD SIMAR

VALENTIN ZELENYUK

SHIRONG ZHAO*

October 2022

Abstract

A simple yet easy to implement method is proposed to further improve the finite sample approximation of the recently developed central limit theorems for aggregates of envelopment estimators. Focusing on the simple mean efficiency, we propose using the bias-corrected individual efficiency estimate to improve the variance estimator. The extensive Monte-Carlo experiments confirm that, for relatively small sample sizes (≤ 100), with both low dimensions and especially for high dimensions, our new method combined with the data sharpening method generally provides better ‘coverage’ (of the true values by the estimated confidence intervals) than the previously developed approaches.

Keywords: Efficiency, Non-parametric Efficiency Estimators, Data Envelopment Analysis, Free Disposal Hull

JEL Classification: C1, C3

*Simar: Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Voie du Roman Pays 20, B1348 Louvain-la-Neuve, Belgium; email: leopold.simar@uclouvain.be.

Zelenyuk: School of Economics and Centre for Efficiency and Productivity Analysis (CEPA), University of Queensland, Colin Clark Building (39), St Lucia, Brisbane, Qld 4072, Australia; email: v.zelenyuk@uq.edu.au.

Zhao (Corresponding Author): School of Finance, Dongbei University of Finance and Economics, Dalian, Liaoning 116025, China; email: shironz@dufe.edu.cn.

1 Introduction

Nonparametric efficiency estimators are widely used to examine the efficiency and productivity of decision making units (banks, schools, hospitals, countries, etc.). Among the most popular of these are the data envelopment analysis (DEA) estimators that assume constant returns to scale (CRS-DEA), variable returns to scale (VRS-DEA) and those that relax the assumption of convexity, such as the free disposal hull (FDH) estimator (Farrell, 1957; Charnes et al., 1978; Banker et al., 1984; Deprins et al., 1984).

The theoretical and statistical properties for these estimators are now well known in the literature (for a guided tour, see Simar and Wilson, 2015). Specifically, Kneip et al. (2015) have established the new CLT results for the mean of DEA-estimated technical efficiency. This discovery has paved the way for the development of CLT results for other related efficiency and productivity measures, such as aggregate efficiency (Simar and Zelenyuk, 2018), conditional efficiency (Daraio et al., 2018), Malmquist indices (Kneip et al., 2021), etc.

Extensive Monte Carlo (MC) experiments in both Kneip et al. (2015) and Simar and Zelenyuk (2018) find that when the sample sizes increase, the empirical coverage of the true values by estimated confidence intervals based on the CLT, approximate the nominal coverage, supporting the CLT results. However, they also show that the CLT does not perform that well for relatively small sample sizes, especially for high dimensions (measured by the number of inputs and outputs). In particular, the estimated confidence intervals based on the CLT tend to under-cover the true values, especially for relatively small sample sizes and high dimensions.

Simar and Zelenyuk (2020) propose a variance correction method by adding the squared bias of mean efficiency into the variance estimator, thus making the variance estimate slightly larger than the case without adding the squared bias (i.e., the original method in Kneip et al., 2015 and Simar and Zelenyuk, 2018). The ~~extensive~~ MC experiments in Simar and Zelenyuk (2020) suggest a better performance of the proposed variance correction method over the original method in Kneip et al. (2015) and Simar and Zelenyuk (2018); however, they also suggest that the coverage of this variance correction method can still be substantially smaller than the nominal coverage. Nguyen et al. (2022) then suggest an alternative way, a variant of “data sharpening”, made through smoothing out the so-called “spurious ones” and

values within its neighborhood (determined based on the convergence rate) of the efficiency estimates. [Nguyen et al. \(2022\)](#) find that combining the data sharpening method and the variance correction method results in a better performance than all the current methods in the literature. In their concluding section, [Nguyen et al. \(2022\)](#) also state that “It is also worth mentioning here that even with the proposed improvements, the empirical coverages of the CLTs for small samples, with a relatively large dimension of inputs and outputs are far from the nominal coverages”. This indicates that there is still some room for improvement.

In this paper, we propose an alternative method for the variance estimator, by using the *bias-corrected individual efficiency* estimate instead of the individual efficiency estimate. Our proposed method does not involve any additional computational cost and retains the original CLT results. Furthermore, we provide theoretical justification for our method. We also provide simulations to illustrate the effectiveness of our method on improving the finite sample approximation of CLT for efficiency estimates based on DEA.

2 Improving the Approximation

Let $\lambda_i = \lambda(X_i, Y_i)$ be the true technical efficiency at a random point $(X_i, Y_i) \in \Psi$, where Ψ is the attainable set in the input-output space, with a data generating process that satisfies regularity conditions in [Kneip et al. \(2015\)](#). Suppose we have already obtained the efficiency estimate $\hat{\lambda}_i$ for each observation using either FDH or VRS-DEA or CRS-DEA estimator, which is consistent and its convergence rate is n^κ , meaning that the order of the error of estimation is $O(n^{-\kappa})$, where κ depends on the estimator used and the dimensions of the problem. In particular, $\kappa = 1/(p + q)$ for the FDH estimator, $\kappa = 2/(p + q + 1)$ for the VRS-DEA estimator and $\kappa = 2/(p + q)$ for the CRS-DEA estimator, where p is the number of inputs and q is the number of outputs. Suppose we also obtain its bias $\hat{B}_i$, estimated using the generalized jackknife bootstrap method of [Kneip et al. \(2015\)](#), where more details can be found.

Let $\mu = E(\lambda(X_i, Y_i))$ and so its sample counterpart is $\bar{\lambda} = n^{-1} \sum_{i=1}^n \lambda_i$. Both of these are unavailable but can be estimated, e.g., with the FDH/DEA-based estimator, which we denote by $\hat{\mu} = n^{-1} \sum_{i=1}^n \hat{\lambda}_i$. This estimator is biased and, as shown in [Kneip et al. \(2015\)](#), $\hat{B} = n^{-1} \sum_{i=1}^n \hat{B}_i$ is a consistent estimate of the bias of $\hat{\mu}$. Furthermore, let $\sigma^2 = \text{Var}(\lambda(X_i, Y_i))$,

then its sample counterpart is given by

$$\bar{\sigma}^2 = \widehat{\text{Var}}(\lambda_i) = n^{-1} \sum_{i=1}^n (\lambda_i - n^{-1} \sum_{i=1}^n \lambda_i)^2. \quad (2.1)$$

To perform statistical inference, it would be desirable to know σ^2 or at least $\bar{\sigma}^2$, yet both of them are unavailable. [Kneip et al. \(2015\)](#) propose the following method to estimate σ^2 ,

$$\hat{\sigma}^2 = \widehat{\text{Var}}(\hat{\lambda}_i) = n^{-1} \sum_{i=1}^n (\hat{\lambda}_i - n^{-1} \sum_{i=1}^n \hat{\lambda}_i)^2, \quad (2.2)$$

i.e., it implies replacing λ_i by $\hat{\lambda}_i$ at every place in (2.1). By noting that $n^{-1} \sum_{i=1}^n \hat{\lambda}_i - \hat{B} = n^{-1} \sum_{i=1}^n (\hat{\lambda}_i - \hat{B}_i)$ is a bias-corrected estimator of μ , [Simar and Zelenyuk \(2020\)](#) propose the following method,

$$\tilde{\sigma}^2 = n^{-1} \sum_{i=1}^n \left[\hat{\lambda}_i - (n^{-1} \sum_{i=1}^n \hat{\lambda}_i - \hat{B}) \right]^2 = \hat{\sigma}^2 + \hat{B}^2. \quad (2.3)$$

In this paper, we consider the question of using the *bias-corrected individual efficiency estimate*, $\hat{\lambda}_i - \hat{B}_i$, to replace λ_i at every place in (2.1). We explain the resulting theoretical properties of this approach as well as possible practical benefits, as evidenced from its Monte Carlo performance relative to other approaches. That is, instead of $\hat{\sigma}^2$ or $\tilde{\sigma}^2$, we propose using

$$\hat{\hat{\sigma}}^2 = \widehat{\text{Var}}(\hat{\lambda}_i - \hat{B}_i) = n^{-1} \sum_{i=1}^n \left[(\hat{\lambda}_i - \hat{B}_i) - n^{-1} \sum_{i=1}^n (\hat{\lambda}_i - \hat{B}_i) \right]^2. \quad (2.4)$$

Compared with (2.2) and (2.3), (2.4) has potential to improve, as (2.4) replaces the $\hat{\lambda}_i$ in (2.1) by its bias-corrected estimate $\hat{\lambda}_i - \hat{B}_i$. To see this, recall that $\sigma^2 = \mu_2 - \mu^2$, where $\mu_2 = E(\lambda_i^2)$ and so we may view σ^2 as a function of the first two raw moments (μ, μ_2) , and hence the Taylor expansion of $\hat{\sigma}^2$ around σ^2 is

$$\hat{\sigma}^2 = \sigma^2 + A_1(\hat{\mu} - \mu) + A_2(\hat{\mu}_2 - \mu_2) + R_{1,n}, \quad (2.5)$$

while the Taylor expansion of $\hat{\hat{\sigma}}^2$ around σ^2 is

$$\hat{\hat{\sigma}}^2 = \sigma^2 + A_1(\hat{\hat{\mu}} - \mu) + A_2(\hat{\hat{\mu}}_2 - \mu_2) + R_{2,n}, \quad (2.6)$$

where $A_1 = -2\mu$, $A_2 = 1$, $\hat{\mu}_2 = n^{-1} \sum_{i=1}^n \lambda_i^2$, $\hat{\hat{\mu}} = \hat{\mu} - \hat{B}$ and $\hat{\hat{\mu}}_2 = n^{-1} \sum_{i=1}^n (\hat{\lambda}_i - \hat{B}_i)^2$, while $R_{1,n}$ and $R_{2,n}$ are the remainders of a smaller order, i.e., $o(n^{-\kappa})$, and can be ignored for our purposes.

According to equation (4.2) in [Kneip et al. \(2015\)](#), the first order error (i.e., the dominating part of the error) is given by $A_2(\hat{\mu}_2 - \mu_2) = o(n^{-\kappa/2})$ in (2.5) or $A_2(\hat{\hat{\mu}}_2 - \mu_2) = o(n^{-\kappa/2})$ in (2.6). Thus, while using the bias-corrected individual efficiency estimate does not improve the first order error (keeping it as $o(n^{-\kappa/2})$), we still are able to improve the second order term. Indeed, the error multiplying A_1 will be reduced from $A_1(\hat{\mu} - \mu) = O(n^{-\kappa})$ in (2.5) to $A_1(\hat{\hat{\mu}} - \mu) = o(n^{-\kappa})$ in (2.6), which is an improvement.

Importantly, note that by using the bias-corrected individual efficiency estimate, we are able to keep the asymptotic properties established by [Kneip et al. \(2015\)](#), including their central limit theorems. This means that for relatively large samples the new method should give very similar results (due to their asymptotic equivalence).

How substantial is this improvement in practice? While asymptotically equivalent, the proposed improvement can be substantial in finite samples, as it is possible that the second order error is larger than the first order error. In other words, for relatively small samples it is possible that the $O(n^{-\kappa})$ term is **substantially** bigger than the $o(n^{-\kappa/2})$ term, and the $o(n^{-\kappa})$ term is **substantially** smaller than $o(n^{-\kappa/2})$ term. The improvement is largely confirmed by our simulations presented in the next Section, although, as in general, it might still be possible to find some peculiar scenarios and/or samples where it is different.

We name the method in (2.4) as the ‘full variance correction’ approach to distinguish it from the (partial) variance correction method in [Simar and Zelenyuk \(2020\)](#). Furthermore, as for [Simar and Zelenyuk \(2020\)](#), our method can also be deployed in a combination with the data sharpening approach proposed by [Nguyen et al. \(2022\)](#), where more details of how the sharpening can be done are provided.

The discussion above also helps to shed more light onto potential further improvements. In particular, one would wish to reduce the first order term, yet at this stage we have not discovered a way to do so and hope this paper will encourage more research in this direction.

3 Monte-Carlo Evidence

We conducted extensive Monte-Carlo experiments to compare the performance of our proposed method with those in [Kneip et al. \(2015\)](#), [Simar and Zelenyuk \(2020\)](#) and [Nguyen et al. \(2022\)](#). Following [Simar and Zelenyuk \(2020\)](#), we choose the Cobb-Douglas production function to simulate the data; **and for each experiment (i.e., for each combination of p , q and**

n) we perform 2,000 trials.¹ To simplify the notation, we denote

- Sol1: the original method in Kneip et al. (2015), i.e., the method in (2.2).
- Sol2: the partial variance correction method in Simar and Zelenyuk (2020), i.e., the method in (2.3).
- Sol3: our full variance correction method, i.e., the method in (2.4).
- Sol4: Sol1 combined with the data sharpening in Nguyen et al. (2022).²
- Sol5: Sol2 combined with the data sharpening in Nguyen et al. (2022).
- Sol6: Sol3 combined with the data sharpening in Nguyen et al. (2022).

The Monte Carlo simulations from Figure 1 confirm the results from Simar and Zelenyuk (2020) and Nguyen et al. (2022) that, for small sample sizes and large dimensions, increasing order of performance lie Sol1, Sol2, Sol4 and Sol5.³ For instance, when $n = 100$, $p = 3$, $q = 1$, the empirical coverage under Sol1, Sol2, Sol4 and Sol5 is 0.694, 0.750, 0.828, and 0.890, respectively; however, they are still significantly far from the nominal coverage of 0.95. Regarding our full variance correction method, Sol3 provides significant and persistent improvement compared to Sol1, Sol2 and Sol4. Further, note that the performance of Sol3 and Sol5 is nearly identical. For instance, for $n = 100$, $p = 3$, $q = 1$, the empirical coverage of Sol3 is 0.867, higher than that for Sol1, Sol2 and Sol4, and is very close to that for Sol5.

When adapting the data sharpening method to our full variance correction approach, the improvements of Sol6 are even more significant and persistent relative to all the other five methods, especially when the sample size is smaller than 200. For instance, for $n = 100$, $p = 3$, $q = 1$, the empirical coverage under Sol6 is 0.954, higher than all the other five

¹ For more details about the simulation process, see the Appendix A.

² The data sharpening method works by smoothing-out the estimated efficiency scores that are close to 1, i.e., those smaller than $1 + \tau$, where $\tau = n^{-\gamma}$ and γ is in the range of $[\kappa/2, \kappa]$. The idea is to use a uniform density with the range $[1, 1 + \tau]$ to approximate the density of the efficiency locally in the τ neighborhood of the frontier. Further, the results from simulations in Nguyen et al. (2022) indicate the level of γ near 0.75κ seems to perform best.

³ For the values of the empirical coverages, see Table A.2 in the separate Appendix A. Further, when $\kappa = 2/5$ for VRS-DEA, Kneip et al. (2015) finds that both Theorem 4.3 and Theorem 4.4 in Kneip et al. (2015) are applicable, while Theorem 4.4 is preferred. Thus, for $p = 3$ and $q = 1$, we only include the results from the preferred method.

methods and very close to the nominal coverage of 0.95. When the sample size increases, the magnitudes of improvements of Sol6 and Sol3 diminish. However, from Figure 1 we see that Sol6 slightly overshoots sometimes, especially when the dimensions are high and the sample size is relatively large. For instance, when $n = 200$, $p = 5$ and $q = 1$, the empirical coverage under Sol6 is 0.999 when the nominal coverage is 0.95. Note that we also observe some overshooting phenomenon for the other methods. It is also worth noting that the error due to such overshooting can be often considered as more preferable than that due to undershooting, if the confidence intervals are of similar lengths (as they are). Indeed, the overshooting implies *not rejecting* the *correct* null hypothesis more often than it should be, i.e., a more conservative approach. Meanwhile, the undershooting or under-covering implies *rejecting* the *correct* null hypothesis more often than should be.

To summarize, our general conclusion is that, when the dimensions are not high (≤ 3), our full variance correction method combined with the data sharpening method (Sol6) provides a better performance for sample sizes (≤ 1000); when the dimensions are high (≥ 4), Sol6 still provides a better performance when the sample size is not large (≤ 100) while the Sol5 remains a good competitor when the sample size is larger. Finally, they give almost identical performances for large samples (e.g., $n = 1000$), as expected due to asymptotic equivalence.⁴

We also use FDH estimators to illustrate the improvements of our proposed method.⁵ The results are summarized in Figure 2.⁶ Figure 2 shows that the improvement of Sol6 relative to others is even more substantial for FDH estimators. For example, when $n = 100$, $p = 3$, $q = 1$ and the nominal coverage is 0.95, the empirical coverage under Sol1, Sol2, Sol3, Sol4, Sol5 and Sol6 is 0.119, 0.246, 0.753, 0.538, 0.750, and 0.981, respectively. Further, when the dimensions are high, the improvement of Sol6 relative to others is more substantial for FDH estimators.

4 Conclusions

The new CLTs for the mean technical efficiency estimated via DEA and FDH developed by Kneip et al. (2015) are useful recent advancements of the statistical theory for efficiency and

⁴ The simulation results for CRS-DEA estimators are summarized in Figure C.1 of the separate appendix C, which also confirm the improved performance of our proposed method.

⁵ We thank an anonymous referee for encouraging exploring this avenue.

⁶ For more details, see the separate Appendix B.

productivity measurements. While confirmed for large samples, for relatively small samples they performed somewhat disappointingly and so further improvements for such small samples are clearly desired. Some of such improvements were made through [Simar and Zelenyuk \(2020\)](#) and [Nguyen et al. \(2022\)](#), yet still for relatively small samples the approximation remained somewhat disappointing (although better). Hence, further improvements are desired and here we propose such an improvement, which can also be combined with the data-sharpening approach of [Nguyen et al. \(2022\)](#).

Specifically, we propose using the bias-corrected individual efficiency estimate to obtain the variance estimator. This newly proposed method involves no additional cost and preserves the CLT results in [Kneip et al. \(2015\)](#). Furthermore, we provide a theoretical justification for the proposed improvement: We show that the improvement corrects the second order of the error in the variance estimate while keeping the first order of the error as in [Kneip et al. \(2015\)](#). While theoretically modest, this improvement can be substantial in finite samples, i.e., in practice, as is evident from our extensive Monte-Carlo experiments. Indeed, for low dimensions or for high dimensions with small sample sizes (≤ 100), our method combined with the data sharpening method in [Nguyen et al. \(2022\)](#) provided better (larger) coverages of the true values by the confidence intervals than the original method in [Kneip et al. \(2015\)](#), the variance correction method in [Simar and Zelenyuk \(2020\)](#) and the data sharpening method in [Nguyen et al. \(2022\)](#). **Our method could be useful for several recent applications using CLT results for envelopment estimators with finite sample sizes, such as [Nguyen and Zelenyuk \(2021\)](#) for public hospitals, [Qiao and Wang \(2021\)](#) for the semiconductor industry, [O’Loughlin and Wilson \(2021\)](#) for US Municipalities, and [Wilson and Zhao \(2022\)](#) for Chinese banks.**

Therefore, the proposed improvement appears to be a useful approach (although of course not a panacea) in the toolbox along with other approaches in the literature. This paper also shows that there is still room for improvement, e.g., to reduce the first order term, which at this stage we have not discovered a way to do and so we hope this paper will encourage further research that hopefully brings more light and insights on this challenging topic.

Finally, all what is developed in this short note can be adapted to other contexts, including the context of aggregate efficiency measures analyzed in [Simar and Zelenyuk \(2018\)](#) and combined together with the data sharpening ideas from [Nguyen et al. \(2022\)](#).

Acknowledgement

Valentin Zelenyuk acknowledges the support from the Australian Research Council (FT170100401) and The University of Queensland. Shirong Zhao acknowledges the support from Dongbei University of Finance and Economics. Feedback from Bao Hoang Nguyen, Evelyn Smart and Zhichao Wang is appreciated. These individuals and organizations are not responsible for the views expressed here.

Disclosure statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Banker, R. D., A. Charnes, and W. W. Cooper (1984), Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science* 30, 1078–1092.
- Charnes, A., W. W. Cooper, and E. Rhodes (1978), Measuring the efficiency of decision making units, *European Journal of Operational Research* 2, 429–444.
- Daraio, C., L. Simar, and P. W. Wilson (2018), Central limit theorems for conditional efficiency measures and tests of the ‘separability condition’ in non-parametric, two-stage models of production, *Econometrics Journal* 21, 170–191.
- Deprins, D., L. Simar, and H. Tulkens (1984), Measuring labor inefficiency in post offices, in M. M. P. Pestieau and H. Tulkens, eds., *The Performance of Public Enterprises: Concepts and Measurements*, Amsterdam: North-Holland, pp. 243–267.
- Farrell, M. J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society A* 120, 253–281.
- Kneip, A., L. Simar, and P. W. Wilson (2015), When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores, *Econometric Theory* 31, 394–422.
- (2021), Inference in dynamic, nonparametric models of production: Central limit theorems for malmquist indices, *Econometric Theory* 37, 537–572. Publisher: Cambridge University Press.
- Nguyen, B. H. and V. Zelenyuk (2021), Aggregate efficiency of industry and its groups: the case of queensland public hospitals, *Empirical Economics* 60, 2795–2836.
- Nguyen, H., L. Simar, and V. Zelenyuk (2022), Data sharpening for improving CLT approximations for DEA-type efficiency estimators, *European Journal of Operational Research*, forthcoming.
- O’Loughlin, C. T. and P. W. Wilson (2021), Benchmarking the performance of US municipalities, *Empirical economics* 60, 2665–2700.
- Qiao, G. and Z.-a. Wang (2021), Vertical integration vs. specialization: a nonparametric conditional efficiency estimate for the global semiconductor industry, *Journal of Productivity Analysis* 56, 139–150.
- Simar, L. and P. W. Wilson (2015), Statistical approaches for non-parametric frontier models: A guided tour, *International Statistical Review* 83, 77–110.
- Simar, L. and V. Zelenyuk (2018), Central limit theorems for aggregate efficiency, *Operations Research* 66, 137–149.
- (2020), Improving finite sample approximation by central limit theorems for estimates from Data Envelopment Analysis, *European Journal of Operational Research* 284, 1002–1015.
- Wilson, P. W. and S. Zhao (2022), Investigating the performance of Chinese banks over 2007–2014, *Annals of Operations Research*, Forthcoming.

Figure 1: Empirical Coverages for the Simple Mean Efficiency using VRS-DEA Estimator

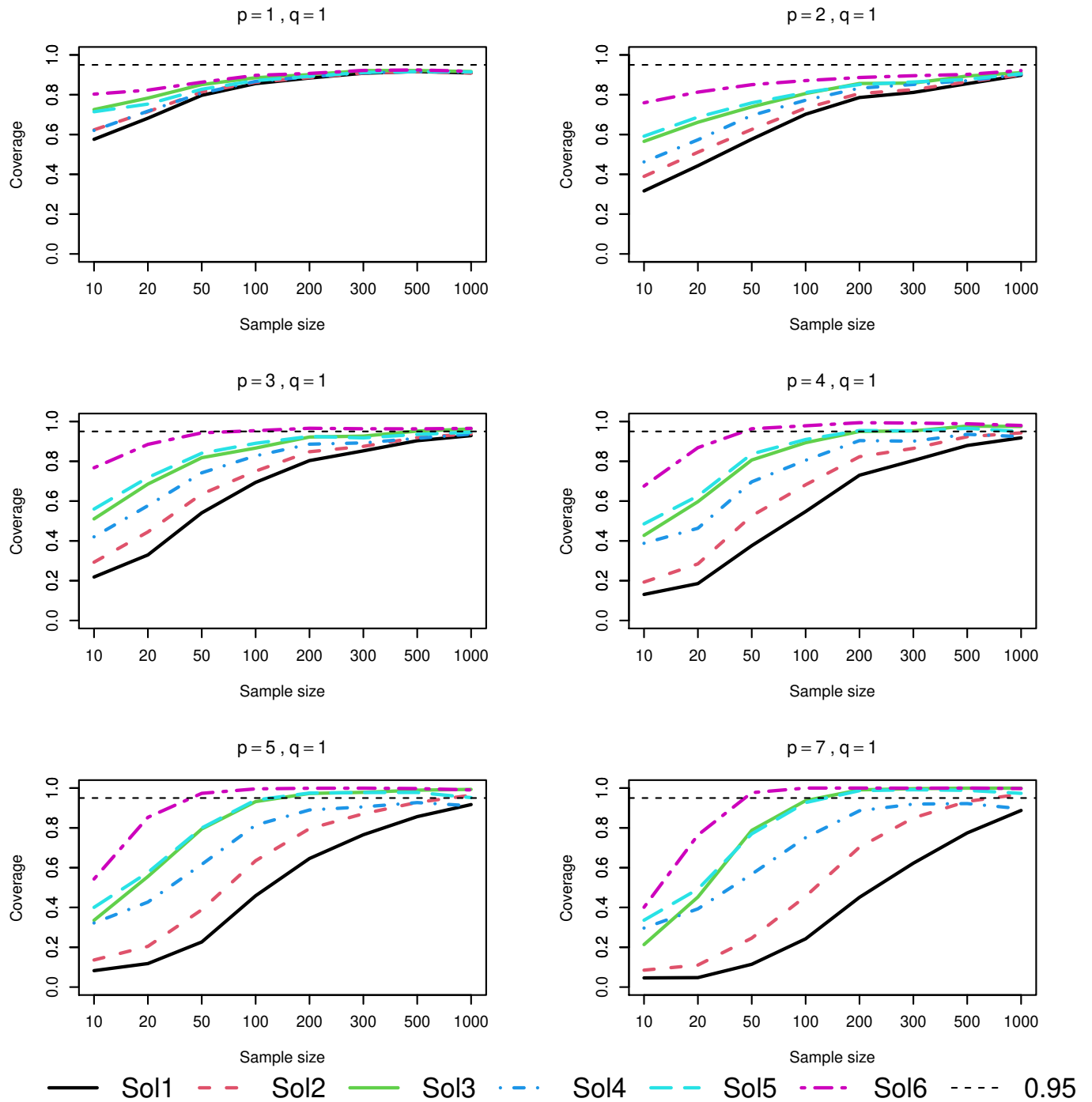
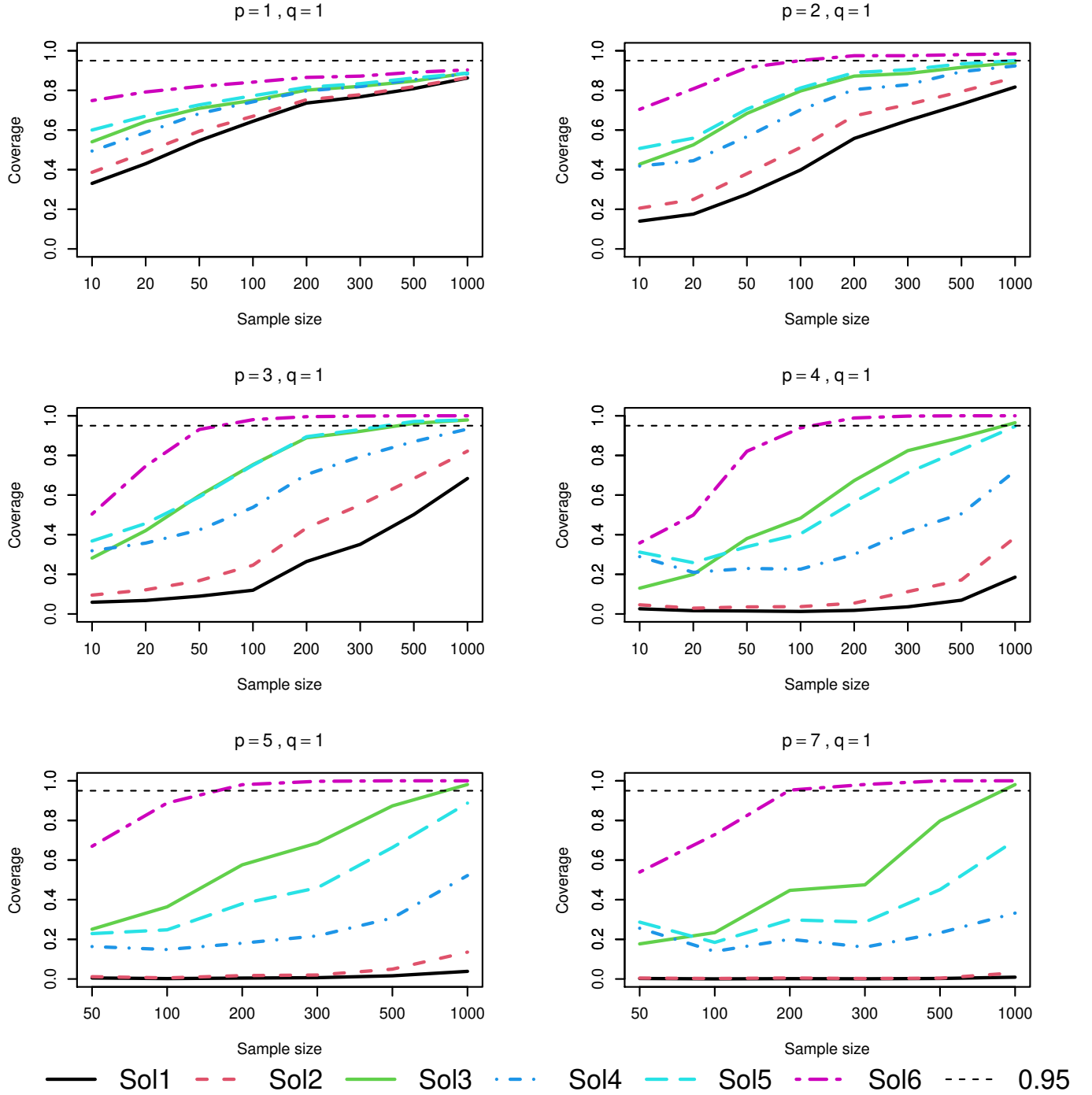


Figure 2: Empirical Coverages for the Simple Mean Efficiency using FDH Estimator



Appendix for “Further Improvements of Finite Sample Approximation of Central Limit Theorems for Envelopment Estimator”

LÉOPOLD SIMAR

VALENTIN ZELENYUK

SHIRONG ZHAO*

October 2022

*Simar: Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Voie du Roman Pays 20, B1348 Louvain-la-Neuve, Belgium; email: leopold.simar@uclouvain.be.

Zelenyuk: School of Economics and Centre for Efficiency and Productivity Analysis (CEPA), University of Queensland, Colin Clark Building (39), St Lucia, Brisbane, Qld 4072, Australia; email: v.zelenyuk@uq.edu.au.

Zhao (Corresponding Author): School of Finance, Dongbei University of Finance and Economics, Dalian, Liaoning 116025, China; email: shironz@dufe.edu.cn.

Valentin Zelenyuk acknowledges the support from the Australian Research Council (FT170100401) and The University of Queensland. Shirong Zhao acknowledges the support from Dongbei University of Finance and Economics. Feedback from Bao Hoang Nguyen, Evelyn Smart and Zhichao Wang is appreciated. These individuals and organizations are not responsible for the views expressed here.

Contents

A	Details and Additional Results from Simulations for VRS-DEA Estimator	1
B	Details and Additional Results from Simulations for FDH Estimator	21
C	Details and Additional Results from Simulations for CRS-DEA Estimator	40

List of Tables

A.1	The Values of β_j	2
A.2	Empirical Coverages for the Simple Mean Efficiency using VRS-DEA Estimator	3
A.3	Simple Mean Efficiency and Bias using VRS-DEA Estimator	9
A.4	Mean Standard Deviation of Estimated Efficiency using VRS-DEA Estimator	15
B.5	Empirical Coverages for the Simple Mean Efficiency using FDH Estimator .	22
B.6	Simple Mean Efficiency and Bias using FDH Estimator	28
B.7	Mean Standard Deviation of Estimated Efficiency using FDH Estimator . . .	34

List of Figures

C.1	Empirical Coverages for the Simple Mean Efficiency using CRS-DEA Estimator	41
-----	--	----

Appendix A Details and Additional Results from Simulations for VRS-DEA Estimator

We conduct extensive Monte-Carlo experiments to compare the performance of our proposed method with those in [Kneip et al. \(2015\)](#), [Simar and Zelenyuk \(2020\)](#) and [Nguyen et al. \(2022\)](#). Following [Simar and Zelenyuk \(2020\)](#), we choose the Cobb-Douglas production function (VRS) to simulate the data. Formally,

$$y_i^\partial(x_i) = \prod_{j=1}^p x_{ji}^{\beta_j}, \quad (\text{A.1})$$

where $1 \times p$ vector $x_i = (x_{1i}, x_{2i}, \dots, x_{pi})$ and $x_{ji} \stackrel{\text{iid}}{\sim} \text{Unif}(0, 1)$, $\forall j \in 1, \dots, p$. In addition, we generate the true inefficiency scores from $\lambda_i \sim |N(0, 1)| + 1$ (i.e., we generate the true inefficiency from the half normal distribution and then shift it up by 1); it can be shown that the simple mean true efficiency equal to $1 + \sqrt{\frac{2}{\pi}}$. The observed output is $y_i = y_i^\partial(x_i)/\lambda_i$, i.e., by projecting $y_i^\partial(x_i)$ from the frontier into the production set. Thus a simulated sample $\mathcal{S}_n = \{x_i, y_i\}_{i=1}^n$ is created. The values of β_j 's are set the same as [Simar and Zelenyuk \(2020\)](#) so that we can directly compare our method with the correction method in [Simar and Zelenyuk \(2020\)](#).

To be more clear, for each experiment (i.e., for each combination of p , q and n) we perform 2,000 trials, drawing on each trial a sample of size n and checking whether the estimated confidence interval covers the simple mean true efficiency. We then take the average of each trial's coverage over 2,000 trials. Finally, we consider 0.90, 0.95 and 0.99 confidence intervals.

Tables [A.1–A.4](#) present additional results not shown in the paper. Table [A.1](#) lists the values of β_j 's used by [Simar and Zelenyuk \(2020\)](#) for each dimension p . Table [A.2](#) presents the coverages of estimated confidence intervals for the simple mean efficiency using VRS-DEA estimator across different sample sizes and dimensions. Table [A.3](#) presents simple mean efficiency estimates and its estimated bias using VRS-DEA estimator, while Table [A.4](#) presents mean standard deviation of estimated efficiency using VRS-DEA estimator.

Table A.1: The Values of β_j

p	1	2	3	4	5	7
β_1	0.4	0.4	0.4	0.4	0.4	0.05
β_2		0.2	0.2	0.2	0.2	0.1
β_3			0.1	0.1	0.1	0.15
β_4				0.15	0.15	0.2
β_5					0.05	0.125
β_6						0.075
β_7						0.025

Table A.2: Empirical Coverages for the Simple Mean Efficiency using VRS-DEA Estimator

Panel A: Without Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol1	Sol2	Sol3	Sol1	Sol2	Sol3	Sol1	Sol2	Sol3
1	1	10	0.506	0.571	0.667	0.576	0.624	0.726	0.674	0.723	0.810
1	1	20	0.607	0.643	0.714	0.682	0.715	0.783	0.787	0.821	0.867
1	1	50	0.713	0.731	0.778	0.797	0.814	0.851	0.893	0.901	0.925
1	1	100	0.785	0.792	0.826	0.856	0.863	0.885	0.937	0.941	0.952
1	1	200	0.812	0.815	0.839	0.884	0.886	0.901	0.957	0.959	0.963
1	1	300	0.838	0.838	0.857	0.908	0.909	0.921	0.971	0.972	0.975
1	1	500	0.845	0.846	0.856	0.916	0.916	0.921	0.977	0.977	0.979
1	1	1000	0.858	0.858	0.866	0.909	0.909	0.916	0.976	0.976	0.978

Panel B: With Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol4	Sol5	Sol6	Sol4	Sol5	Sol6	Sol4	Sol5	Sol6
1	1	10	0.550	0.640	0.752	0.621	0.715	0.803	0.739	0.809	0.878
1	1	20	0.646	0.692	0.757	0.718	0.753	0.823	0.817	0.852	0.892
1	1	50	0.734	0.750	0.796	0.808	0.828	0.863	0.902	0.916	0.943
1	1	100	0.803	0.807	0.826	0.867	0.871	0.897	0.943	0.947	0.958
1	1	200	0.817	0.820	0.841	0.890	0.892	0.907	0.958	0.959	0.967
1	1	300	0.847	0.849	0.864	0.910	0.912	0.922	0.973	0.974	0.978
1	1	500	0.844	0.847	0.856	0.916	0.916	0.924	0.978	0.979	0.980
1	1	1000	0.860	0.860	0.868	0.914	0.914	0.917	0.976	0.976	0.979

Table A.2: Empirical Coverages for the Simple Mean Efficiency using VRS-DEA Estimator
(continued)

Panel A: Without Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol1	Sol2	Sol3	Sol1	Sol2	Sol3	Sol1	Sol2	Sol3
2	1	10	0.273	0.339	0.503	0.317	0.390	0.566	0.398	0.484	0.667
2	1	20	0.374	0.442	0.584	0.443	0.511	0.661	0.556	0.622	0.783
2	1	50	0.508	0.550	0.656	0.577	0.625	0.739	0.721	0.773	0.871
2	1	100	0.629	0.661	0.722	0.702	0.733	0.806	0.827	0.856	0.911
2	1	200	0.709	0.729	0.778	0.786	0.806	0.857	0.897	0.907	0.937
2	1	300	0.738	0.747	0.788	0.811	0.826	0.859	0.914	0.920	0.947
2	1	500	0.768	0.779	0.820	0.856	0.865	0.894	0.942	0.947	0.961
2	1	1000	0.828	0.833	0.851	0.897	0.900	0.911	0.956	0.960	0.969

Panel B: With Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol4	Sol5	Sol6	Sol4	Sol5	Sol6	Sol4	Sol5	Sol6
2	1	10	0.399	0.531	0.703	0.463	0.592	0.759	0.564	0.687	0.845
2	1	20	0.503	0.606	0.746	0.574	0.688	0.813	0.702	0.790	0.896
2	1	50	0.607	0.671	0.775	0.696	0.759	0.851	0.826	0.868	0.931
2	1	100	0.684	0.723	0.803	0.772	0.811	0.871	0.888	0.913	0.953
2	1	200	0.754	0.773	0.828	0.833	0.852	0.886	0.926	0.940	0.962
2	1	300	0.770	0.786	0.829	0.852	0.864	0.895	0.935	0.943	0.962
2	1	500	0.792	0.803	0.829	0.871	0.877	0.902	0.953	0.958	0.973
2	1	1000	0.844	0.849	0.867	0.902	0.905	0.921	0.968	0.969	0.975

Table A.2: Empirical Coverages for the Simple Mean Efficiency using VRS-DEA Estimator
(continued)

Panel A: Without Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol1	Sol2	Sol3	Sol1	Sol2	Sol3	Sol1	Sol2	Sol3
3	1	10	0.177	0.254	0.450	0.218	0.293	0.510	0.277	0.367	0.625
3	1	20	0.264	0.371	0.598	0.330	0.445	0.685	0.441	0.560	0.808
3	1	50	0.458	0.545	0.736	0.541	0.635	0.818	0.695	0.783	0.926
3	1	100	0.612	0.679	0.797	0.694	0.750	0.867	0.822	0.878	0.953
3	1	200	0.724	0.771	0.855	0.803	0.848	0.922	0.916	0.947	0.981
3	1	300	0.766	0.807	0.872	0.853	0.875	0.926	0.939	0.960	0.984
3	1	500	0.837	0.857	0.910	0.903	0.919	0.952	0.968	0.978	0.991
3	1	1000	0.867	0.879	0.919	0.929	0.938	0.963	0.985	0.989	0.995

Panel B: With Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol4	Sol5	Sol6	Sol4	Sol5	Sol6	Sol4	Sol5	Sol6
3	1	10	0.361	0.498	0.717	0.420	0.560	0.767	0.531	0.655	0.833
3	1	20	0.498	0.648	0.834	0.577	0.719	0.885	0.688	0.812	0.947
3	1	50	0.665	0.776	0.896	0.742	0.841	0.943	0.852	0.920	0.980
3	1	100	0.744	0.824	0.914	0.828	0.890	0.954	0.922	0.956	0.989
3	1	200	0.805	0.864	0.932	0.886	0.925	0.966	0.959	0.982	0.994
3	1	300	0.823	0.854	0.915	0.892	0.917	0.964	0.967	0.978	0.994
3	1	500	0.859	0.878	0.926	0.915	0.935	0.963	0.979	0.985	0.997
3	1	1000	0.870	0.886	0.925	0.937	0.947	0.966	0.983	0.988	0.996

Table A.2: Empirical Coverages for the Simple Mean Efficiency using VRS-DEA Estimator
(continued)

Panel A: Without Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol1	Sol2	Sol3	Sol1	Sol2	Sol3	Sol1	Sol2	Sol3
4	1	10	0.108	0.166	0.366	0.131	0.193	0.427	0.171	0.266	0.517
4	1	20	0.148	0.234	0.504	0.185	0.284	0.597	0.257	0.384	0.735
4	1	50	0.306	0.425	0.707	0.376	0.526	0.806	0.521	0.687	0.925
4	1	100	0.451	0.573	0.800	0.547	0.682	0.893	0.721	0.837	0.968
4	1	200	0.634	0.728	0.875	0.730	0.823	0.950	0.874	0.943	0.994
4	1	300	0.713	0.791	0.893	0.804	0.865	0.953	0.913	0.957	0.995
4	1	500	0.799	0.849	0.931	0.879	0.923	0.977	0.967	0.984	0.998
4	1	1000	0.851	0.883	0.941	0.918	0.943	0.975	0.980	0.987	0.998

Panel B: With Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol4	Sol5	Sol6	Sol4	Sol5	Sol6	Sol4	Sol5	Sol6
4	1	10	0.319	0.439	0.639	0.388	0.485	0.675	0.482	0.578	0.731
4	1	20	0.407	0.554	0.805	0.464	0.625	0.869	0.586	0.722	0.933
4	1	50	0.608	0.771	0.928	0.697	0.837	0.964	0.815	0.914	0.995
4	1	100	0.717	0.850	0.954	0.805	0.909	0.979	0.910	0.969	0.997
4	1	200	0.822	0.911	0.977	0.904	0.955	0.994	0.970	0.995	1.000
4	1	300	0.835	0.900	0.968	0.901	0.953	0.992	0.972	0.993	1.000
4	1	500	0.868	0.928	0.971	0.935	0.966	0.989	0.982	0.992	0.999
4	1	1000	0.870	0.903	0.949	0.924	0.950	0.980	0.980	0.992	0.999

Table A.2: Empirical Coverages for the Simple Mean Efficiency using VRS-DEA Estimator
(continued)

Panel A: Without Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol1	Sol2	Sol3	Sol1	Sol2	Sol3	Sol1	Sol2	Sol3
5	1	10	0.065	0.108	0.287	0.083	0.136	0.336	0.108	0.176	0.402
5	1	20	0.092	0.165	0.464	0.118	0.204	0.555	0.176	0.294	0.685
5	1	50	0.174	0.293	0.682	0.227	0.389	0.794	0.349	0.556	0.912
5	1	100	0.357	0.515	0.843	0.459	0.635	0.932	0.625	0.819	0.989
5	1	200	0.544	0.687	0.912	0.646	0.796	0.973	0.822	0.933	0.997
5	1	300	0.671	0.784	0.930	0.765	0.871	0.979	0.894	0.967	0.999
5	1	500	0.776	0.861	0.963	0.857	0.928	0.990	0.961	0.989	1.000
5	1	1000	0.841	0.904	0.975	0.917	0.964	0.994	0.984	0.996	1.000

Panel B: With Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol4	Sol5	Sol6	Sol4	Sol5	Sol6	Sol4	Sol5	Sol6
5	1	10	0.275	0.360	0.509	0.322	0.401	0.543	0.409	0.472	0.593
5	1	20	0.350	0.517	0.802	0.427	0.576	0.853	0.548	0.685	0.911
5	1	50	0.536	0.732	0.945	0.617	0.799	0.974	0.764	0.895	0.994
5	1	100	0.724	0.891	0.987	0.814	0.940	0.996	0.922	0.984	1.000
5	1	200	0.823	0.931	0.993	0.890	0.975	0.999	0.970	0.997	1.000
5	1	300	0.838	0.938	0.995	0.905	0.978	1.000	0.977	0.999	1.000
5	1	500	0.870	0.939	0.990	0.927	0.978	0.998	0.982	0.997	1.000
5	1	1000	0.846	0.910	0.963	0.911	0.952	0.991	0.971	0.995	1.000

Table A.2: Empirical Coverages for the Simple Mean Efficiency using VRS-DEA Estimator
(continued)

Panel A: Without Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol1	Sol2	Sol3	Sol1	Sol2	Sol3	Sol1	Sol2	Sol3
7	1	10	0.036	0.072	0.190	0.046	0.085	0.213	0.064	0.109	0.251
7	1	20	0.035	0.084	0.382	0.048	0.111	0.453	0.080	0.175	0.572
7	1	50	0.079	0.184	0.674	0.115	0.246	0.787	0.194	0.391	0.917
7	1	100	0.178	0.350	0.862	0.242	0.456	0.937	0.378	0.663	0.990
7	1	200	0.344	0.575	0.947	0.451	0.703	0.991	0.646	0.887	1.000
7	1	300	0.499	0.739	0.980	0.623	0.850	0.997	0.815	0.973	1.000
7	1	500	0.670	0.859	0.992	0.774	0.933	1.000	0.925	0.992	1.000
7	1	1000	0.813	0.920	0.995	0.887	0.970	1.000	0.971	0.998	1.000

Panel B: With Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol4	Sol5	Sol6	Sol4	Sol5	Sol6	Sol4	Sol5	Sol6
7	1	10	0.240	0.288	0.359	0.296	0.336	0.401	0.441	0.466	0.506
7	1	20	0.323	0.435	0.719	0.393	0.492	0.764	0.518	0.604	0.820
7	1	50	0.492	0.689	0.957	0.568	0.770	0.977	0.724	0.878	0.994
7	1	100	0.668	0.874	0.999	0.751	0.927	1.000	0.883	0.982	1.000
7	1	200	0.810	0.962	1.000	0.886	0.987	1.000	0.963	0.999	1.000
7	1	300	0.863	0.976	1.000	0.919	0.992	1.000	0.977	1.000	1.000
7	1	500	0.873	0.972	0.999	0.922	0.989	1.000	0.979	1.000	1.000
7	1	1000	0.838	0.944	0.989	0.896	0.974	0.998	0.964	0.994	1.000

Table A.3: Simple Mean Efficiency and Bias using VRS-DEA Estimator

Panel A: Without Data Sharpening

p	q	n	$\hat{\mu}_n$	$\hat{B}_n$	$\hat{\mu}_n - \hat{B}_n$	μ_s	μ
1	1	10	1.3842	-0.2451	1.6293	1.8065	1.7979
1	1	20	1.5007	-0.2025	1.7032	1.7965	1.7979
1	1	50	1.6200	-0.1443	1.7642	1.8014	1.7979
1	1	100	1.6745	-0.1018	1.7762	1.7956	1.7979
1	1	200	1.7183	-0.0709	1.7892	1.7982	1.7979
1	1	300	1.7345	-0.0563	1.7908	1.7969	1.7979
1	1	500	1.7525	-0.0421	1.7946	1.7981	1.7979
1	1	1000	1.7682	-0.0276	1.7958	1.7974	1.7979

Panel B: With Data Sharpening

p	q	n	$\hat{\mu}_n$	$\hat{B}_n$	$\hat{\mu}_n - \hat{B}_n$	μ_s	μ
1	1	10	1.4806	-0.2508	1.7314	1.8065	1.7979
1	1	20	1.5475	-0.2045	1.7520	1.7965	1.7979
1	1	50	1.6368	-0.1447	1.7815	1.8014	1.7979
1	1	100	1.6824	-0.1019	1.7843	1.7956	1.7979
1	1	200	1.7220	-0.0710	1.7929	1.7982	1.7979
1	1	300	1.7368	-0.0564	1.7932	1.7969	1.7979
1	1	500	1.7539	-0.0421	1.7960	1.7981	1.7979
1	1	1000	1.7688	-0.0276	1.7964	1.7974	1.7979

NOTE: $\hat{\mu}_n$: Mean efficiency constructed using n observations; $\hat{B}_n$: Bias constructed using n observations; μ_s : Sample mean of true efficiency; μ : Population mean of true efficiency.

Table A.3: Simple Mean Efficiency and Bias using VRS-DEA Estimator (continued)

Panel A: Without Data Sharpening

p	q	n	$\hat{\mu}_n$	$\hat{B}_n$	$\hat{\mu}_n - \hat{B}_n$	μ_s	μ
2	1	10	1.2178	-0.2600	1.4778	1.8008	1.7979
2	1	20	1.3300	-0.2686	1.5986	1.8005	1.7979
2	1	50	1.4645	-0.2365	1.7010	1.7980	1.7979
2	1	100	1.5489	-0.1985	1.7474	1.7992	1.7979
2	1	200	1.6127	-0.1578	1.7705	1.7985	1.7979
2	1	300	1.6424	-0.1358	1.7782	1.7969	1.7979
2	1	500	1.6773	-0.1116	1.7889	1.7991	1.7979
2	1	1000	1.7109	-0.0825	1.7935	1.7977	1.7979

Panel B: With Data Sharpening

p	q	n	$\hat{\mu}_n$	$\hat{B}_n$	$\hat{\mu}_n - \hat{B}_n$	μ_s	μ
2	1	10	1.3961	-0.2752	1.6713	1.8008	1.7979
2	1	20	1.4393	-0.2765	1.7158	1.8005	1.7979
2	1	50	1.5172	-0.2395	1.7567	1.7980	1.7979
2	1	100	1.5783	-0.1997	1.7780	1.7992	1.7979
2	1	200	1.6291	-0.1584	1.7875	1.7985	1.7979
2	1	300	1.6540	-0.1362	1.7902	1.7969	1.7979
2	1	500	1.6849	-0.1118	1.7967	1.7991	1.7979
2	1	1000	1.7152	-0.0826	1.7977	1.7977	1.7979

NOTE: $\hat{\mu}_n$: Mean efficiency constructed using n observations; $\hat{B}_n$: Bias constructed using n observations; μ_s : Sample mean of true efficiency; μ : Population mean of true efficiency.

Table A.3: Simple Mean Efficiency and Bias using VRS-DEA Estimator (continued)

Panel A: Without Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
3	1	10	1.1287	-0.2303	1.3590	1.8022	1.7979
3	1	20	1.2165	-0.2841	1.5005	1.7961	1.7979
3	1	50	1.3423	-0.2985	1.6408	1.8005	1.7979
3	1	100	1.4341	-0.2723	1.7064	1.7964	1.7979
3	1	200	1.5097	-0.2393	1.7490	1.7970	1.7979
3	1	300	1.5493	-0.2162	1.7655	1.7972	1.7979
3	1	500	1.5929	-0.1885	1.7814	1.7990	1.7979
3	1	1000	1.6388	-0.1516	1.7904	1.7978	1.7979

Panel B: With Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
3	1	10	1.3608	-0.2529	1.6137	1.8022	1.7979
3	1	20	1.3858	-0.3004	1.6863	1.7961	1.7979
3	1	50	1.4414	-0.3069	1.7483	1.8005	1.7979
3	1	100	1.4976	-0.2769	1.7744	1.7964	1.7979
3	1	200	1.5494	-0.2417	1.7911	1.7970	1.7979
3	1	300	1.5795	-0.2178	1.7973	1.7972	1.7979
3	1	500	1.6140	-0.1895	1.8035	1.7990	1.7979
3	1	1000	1.6520	-0.1521	1.8040	1.7978	1.7979

NOTE: $\hat{\mu}_{n_\kappa}$: Mean efficiency constructed using n_κ observations where $n_\kappa = \lfloor n^{2\kappa} \rfloor$; $\hat{B}_n$: Bias constructed using n observations; μ_s : Sample mean of true efficiency; μ : Population mean of true efficiency.

Table A.3: Simple Mean Efficiency and Bias using VRS-DEA Estimator (continued)

Panel A: Without Data Sharpening							
p	q	n	$\hat{\mu}_n$	$\hat{B}_n$	$\hat{\mu}_n - \hat{B}_n$	μ_s	μ
4	1	10	1.0681	-0.1741	1.2421	1.8001	1.7979
4	1	20	1.1335	-0.2490	1.3825	1.7939	1.7979
4	1	50	1.2414	-0.3119	1.5533	1.8033	1.7979
4	1	100	1.3202	-0.3191	1.6393	1.7972	1.7979
4	1	200	1.3987	-0.3046	1.7033	1.7990	1.7979
4	1	300	1.4424	-0.2894	1.7318	1.7978	1.7979
4	1	500	1.4908	-0.2657	1.7565	1.7984	1.7979
4	1	1000	1.5498	-0.2289	1.7787	1.7973	1.7979

Panel B: With Data Sharpening							
p	q	n	$\hat{\mu}_n$	$\hat{B}_n$	$\hat{\mu}_n - \hat{B}_n$	μ_s	μ
4	1	10	1.3464	-0.1989	1.5453	1.8001	1.7979
4	1	20	1.3551	-0.2724	1.6275	1.7939	1.7979
4	1	50	1.3900	-0.3277	1.7176	1.8033	1.7979
4	1	100	1.4265	-0.3296	1.7562	1.7972	1.7979
4	1	200	1.4719	-0.3111	1.7830	1.7990	1.7979
4	1	300	1.5004	-0.2942	1.7946	1.7978	1.7979
4	1	500	1.5345	-0.2689	1.8035	1.7984	1.7979
4	1	1000	1.5789	-0.2307	1.8096	1.7973	1.7979

NOTE: $\hat{\mu}_{n_\kappa}$: Mean efficiency constructed using n_κ observations where $n_\kappa = \lfloor n^{2\kappa} \rfloor$; $\hat{B}_n$: Bias constructed using n observations; μ_s : Sample mean of true efficiency; μ : Population mean of true efficiency.

Table A.3: Simple Mean Efficiency and Bias using VRS-DEA Estimator (continued)

Panel A: Without Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
5	1	10	1.0382	-0.1248	1.1630	1.7995	1.7979
5	1	20	1.0892	-0.2140	1.3033	1.7972	1.7979
5	1	50	1.1690	-0.2988	1.4678	1.7982	1.7979
5	1	100	1.2462	-0.3383	1.5845	1.7972	1.7979
5	1	200	1.3193	-0.3417	1.6610	1.7960	1.7979
5	1	300	1.3640	-0.3385	1.7025	1.7981	1.7979
5	1	500	1.4126	-0.3229	1.7354	1.7974	1.7979
5	1	1000	1.4792	-0.2931	1.7724	1.7981	1.7979

Panel B: With Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
5	1	10	1.3450	-0.1475	1.4925	1.7995	1.7979
5	1	20	1.3419	-0.2411	1.5830	1.7972	1.7979
5	1	50	1.3601	-0.3219	1.6819	1.7982	1.7979
5	1	100	1.3928	-0.3560	1.7488	1.7972	1.7979
5	1	200	1.4269	-0.3540	1.7809	1.7960	1.7979
5	1	300	1.4539	-0.3482	1.8022	1.7981	1.7979
5	1	500	1.4829	-0.3299	1.8127	1.7974	1.7979
5	1	1000	1.5294	-0.2975	1.8269	1.7981	1.7979

NOTE: $\hat{\mu}_{n_\kappa}$: Mean efficiency constructed using n_κ observations where $n_\kappa = \lfloor n^{2\kappa} \rfloor$; $\hat{B}_n$: Bias constructed using n observations; μ_s : Sample mean of true efficiency; μ : Population mean of true efficiency.

Table A.3: Simple Mean Efficiency and Bias using VRS-DEA Estimator (continued)

Panel A: Without Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
7	1	10	1.0159	-0.0717	1.0877	1.8017	1.7979
7	1	20	1.0397	-0.1448	1.1844	1.8001	1.7979
7	1	50	1.0964	-0.2558	1.3522	1.7971	1.7979
7	1	100	1.1496	-0.3201	1.4696	1.7950	1.7979
7	1	200	1.2056	-0.3641	1.5697	1.7974	1.7979
7	1	300	1.2482	-0.3809	1.6291	1.7980	1.7979
7	1	500	1.2909	-0.3877	1.6786	1.7976	1.7979
7	1	1000	1.3542	-0.3807	1.7349	1.7972	1.7979

Panel B: With Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
7	1	10	1.3541	-0.0884	1.4425	1.8017	1.7979
7	1	20	1.3448	-0.1718	1.5167	1.8001	1.7979
7	1	50	1.3494	-0.2884	1.6377	1.7971	1.7979
7	1	100	1.3619	-0.3510	1.7129	1.7950	1.7979
7	1	200	1.3799	-0.3902	1.7702	1.7974	1.7979
7	1	300	1.4003	-0.4037	1.8040	1.7980	1.7979
7	1	500	1.4212	-0.4065	1.8277	1.7976	1.7979
7	1	1000	1.4550	-0.3944	1.8494	1.7972	1.7979

NOTE: $\hat{\mu}_{n_\kappa}$: Mean efficiency constructed using n_κ observations where $n_\kappa = \lfloor n^{2\kappa} \rfloor$; $\hat{B}_n$: Bias constructed using n observation; μ_s : Sample mean of true efficiency; μ : Population mean of true efficiency.

Table A.4: Mean Standard Deviation of Estimated Efficiency using VRS-DEA Estimator

p	q	n	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	σ_s	σ
1	1	10	0.4459	0.5154	0.6867	0.3899	0.4712	0.6268	0.5841	0.6028
1	1	20	0.5052	0.5485	0.6576	0.4671	0.5143	0.6165	0.5941	0.6028
1	1	50	0.5510	0.5710	0.6345	0.5340	0.5548	0.6164	0.6007	0.6028
1	1	100	0.5678	0.5774	0.6186	0.5591	0.5689	0.6095	0.6005	0.6028
1	1	200	0.5814	0.5859	0.6130	0.5771	0.5816	0.6085	0.6029	0.6028
1	1	300	0.5854	0.5883	0.6091	0.5826	0.5854	0.6062	0.6022	0.6028
1	1	500	0.5900	0.5916	0.6070	0.5884	0.5899	0.6053	0.6026	0.6028
1	1	1000	0.5942	0.5949	0.6047	0.5934	0.5941	0.6039	0.6024	0.6028

NOTE: σ_s : Sample standard deviation of true efficiency; σ : Population standard deviation of true efficiency.

Table A.4: Mean Standard Deviation of Estimated Efficiency using VRS-DEA Estimator
(continued)

p	q	n	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	σ_s	σ
2	1	10	0.3354	0.4310	0.7017	0.2994	0.4146	0.6638	0.5794	0.6028
2	1	20	0.4282	0.5101	0.7229	0.3731	0.4696	0.6639	0.5956	0.6028
2	1	50	0.4968	0.5525	0.6895	0.4572	0.5185	0.6465	0.5976	0.6028
2	1	100	0.5309	0.5678	0.6664	0.5048	0.5440	0.6381	0.6011	0.6028
2	1	200	0.5520	0.5746	0.6445	0.5357	0.5591	0.6270	0.6022	0.6028
2	1	300	0.5605	0.5770	0.6345	0.5483	0.5652	0.6216	0.6018	0.6028
2	1	500	0.5704	0.5813	0.6260	0.5621	0.5732	0.6172	0.6028	0.6028
2	1	1000	0.5794	0.5852	0.6168	0.5745	0.5804	0.6117	0.6026	0.6028

NOTE: σ_s : Sample standard deviation of true efficiency; σ : Population standard deviation of true efficiency.

Table A.4: Mean Standard Deviation of Estimated Efficiency using VRS-DEA Estimator
(continued)

p	q	n	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	σ_s	σ
3	1	10	0.2455	0.3423	0.6674	0.2584	0.3712	0.6836	0.5861	0.6028
3	1	20	0.3474	0.4538	0.7551	0.3057	0.4341	0.7129	0.5916	0.6028
3	1	50	0.4425	0.5366	0.7622	0.3898	0.4992	0.7067	0.6012	0.6028
3	1	100	0.4863	0.5587	0.7269	0.4426	0.5235	0.6800	0.5998	0.6028
3	1	200	0.5179	0.5712	0.6977	0.4855	0.5430	0.6627	0.6008	0.6028
3	1	300	0.5324	0.5750	0.6802	0.5058	0.5512	0.6515	0.6020	0.6028
3	1	500	0.5464	0.5782	0.6633	0.5262	0.5596	0.6416	0.6028	0.6028
3	1	1000	0.5599	0.5802	0.6436	0.5464	0.5673	0.6292	0.6024	0.6028

NOTE: σ_s : Sample standard deviation of true efficiency; σ : Population standard deviation of true efficiency.

Table A.4: Mean Standard Deviation of Estimated Efficiency using VRS-DEA Estimator
(continued)

p	q	n	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	σ_s	σ
4	1	10	0.1608	0.2407	0.5512	0.2320	0.3203	0.6327	0.5823	0.6028
4	1	20	0.2620	0.3658	0.7257	0.2582	0.3808	0.7273	0.5888	0.6028
4	1	50	0.3733	0.4889	0.7979	0.3247	0.4640	0.7502	0.6010	0.6028
4	1	100	0.4295	0.5367	0.7863	0.3772	0.5028	0.7325	0.6007	0.6028
4	1	200	0.4710	0.5617	0.7556	0.4246	0.5272	0.7064	0.6017	0.6028
4	1	300	0.4913	0.5708	0.7385	0.4502	0.5384	0.6945	0.6029	0.6028
4	1	500	0.5122	0.5773	0.7155	0.4779	0.5487	0.6785	0.6032	0.6028
4	1	1000	0.5320	0.5793	0.6854	0.5063	0.5565	0.6576	0.6023	0.6028

NOTE: σ_s : Sample standard deviation of true efficiency; σ : Population standard deviation of true efficiency.

Table A.4: Mean Standard Deviation of Estimated Efficiency using VRS-DEA Estimator
(continued)

p	q	n	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	σ_s	σ
5	1	10	0.1012	0.1626	0.4212	0.2171	0.2817	0.5528	0.5883	0.6028
5	1	20	0.2060	0.3014	0.6920	0.2383	0.3461	0.7335	0.5918	0.6028
5	1	50	0.3147	0.4368	0.8146	0.2846	0.4326	0.7918	0.5993	0.6028
5	1	100	0.3809	0.5111	0.8382	0.3314	0.4881	0.7923	0.5994	0.6028
5	1	200	0.4302	0.5503	0.8167	0.3774	0.5185	0.7643	0.6015	0.6028
5	1	300	0.4552	0.5679	0.8024	0.4047	0.5346	0.7509	0.6024	0.6028
5	1	500	0.4791	0.5781	0.7757	0.4332	0.5449	0.7278	0.6020	0.6028
5	1	1000	0.5068	0.5856	0.7412	0.4689	0.5556	0.7010	0.6028	0.6028

NOTE: σ_s : Sample standard deviation of true efficiency; σ : Population standard deviation of true efficiency.

Table A.4: Mean Standard Deviation of Estimated Efficiency using VRS-DEA Estimator
(continued)

p	q	n	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	σ_s	σ
7	1	10	0.0475	0.0869	0.2562	0.2128	0.2524	0.4396	0.5841	0.6028
7	1	20	0.1246	0.1944	0.5560	0.2212	0.2925	0.6692	0.5900	0.6028
7	1	50	0.2328	0.3490	0.8167	0.2492	0.3847	0.8558	0.5990	0.6028
7	1	100	0.2985	0.4396	0.8915	0.2778	0.4495	0.8913	0.5987	0.6028
7	1	200	0.3568	0.5109	0.9111	0.3140	0.5019	0.8827	0.6006	0.6028
7	1	300	0.3870	0.5437	0.9157	0.3369	0.5265	0.8765	0.6032	0.6028
7	1	500	0.4175	0.5702	0.8972	0.3635	0.5457	0.8504	0.6017	0.6028
7	1	1000	0.4536	0.5924	0.8615	0.4001	0.5620	0.8111	0.6018	0.6028

NOTE: σ_s : Sample standard deviation of true efficiency; σ : Population standard deviation of true efficiency.

Appendix B Details and Additional Results from Simulations for FDH Estimator

In this appendix, we conduct extensive Monte-Carlo experiments to illustrate the performance of our proposed method for FDH estimator. The technology and data generation process are the same as for VRS-DEA estimator presented in Appendix A. Instead of using VRS-DEA estimator, here we use FDH estimator. Note that in this situation, according to [Kneip et al. \(2015\)](#), FDH estimator is also consistent but with a slower convergence rate compared to VRS-DEA estimator.

Tables [B.5–B.7](#) present additional results not shown in the paper. Table [B.5](#) presents the coverages of estimated confidence intervals for the simple mean efficiency using FDH estimator across different sample sizes and dimensions. Table [B.6](#) presents simple mean efficiency estimates and bias using FDH estimator, while Table [B.7](#) presents mean standard deviation of estimated efficiency using FDH estimator.

Table B.5: Empirical Coverages for the Simple Mean Efficiency using FDH Estimator

Panel A: Without Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol1	Sol2	Sol3	Sol1	Sol2	Sol3	Sol1	Sol2	Sol3
1	1	10	0.286	0.338	0.482	0.331	0.387	0.540	0.414	0.481	0.651
1	1	20	0.369	0.416	0.551	0.430	0.488	0.642	0.553	0.619	0.773
1	1	50	0.460	0.502	0.614	0.546	0.594	0.710	0.690	0.736	0.839
1	1	100	0.546	0.583	0.665	0.644	0.669	0.750	0.783	0.808	0.877
1	1	200	0.649	0.664	0.726	0.736	0.753	0.801	0.855	0.872	0.916
1	1	300	0.680	0.694	0.742	0.767	0.778	0.821	0.887	0.895	0.923
1	1	500	0.724	0.734	0.767	0.809	0.819	0.847	0.920	0.926	0.947
1	1	1000	0.782	0.789	0.810	0.863	0.865	0.888	0.948	0.948	0.961

Panel B: With Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol4	Sol5	Sol6	Sol4	Sol5	Sol6	Sol4	Sol5	Sol6
1	1	10	0.430	0.535	0.683	0.494	0.600	0.749	0.600	0.695	0.835
1	1	20	0.519	0.599	0.727	0.588	0.670	0.792	0.702	0.771	0.877
1	1	50	0.599	0.652	0.750	0.683	0.727	0.820	0.796	0.838	0.905
1	1	100	0.645	0.679	0.766	0.743	0.773	0.842	0.855	0.881	0.935
1	1	200	0.724	0.742	0.795	0.797	0.816	0.866	0.910	0.922	0.954
1	1	300	0.737	0.754	0.798	0.820	0.834	0.872	0.917	0.927	0.953
1	1	500	0.773	0.781	0.819	0.855	0.863	0.891	0.946	0.948	0.962
1	1	1000	0.811	0.817	0.844	0.883	0.887	0.903	0.960	0.962	0.972

Table B.5: Empirical Coverages for the Simple Mean Efficiency using FDH Estimator
(continued)

Panel A: Without Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol1	Sol2	Sol3	Sol1	Sol2	Sol3	Sol1	Sol2	Sol3
2	1	10	0.117	0.169	0.368	0.140	0.205	0.427	0.204	0.284	0.517
2	1	20	0.139	0.196	0.432	0.175	0.249	0.525	0.252	0.345	0.667
2	1	50	0.216	0.300	0.562	0.276	0.380	0.683	0.401	0.535	0.849
2	1	100	0.321	0.408	0.683	0.399	0.512	0.797	0.575	0.716	0.933
2	1	200	0.447	0.543	0.782	0.557	0.671	0.872	0.758	0.846	0.975
2	1	300	0.538	0.622	0.780	0.648	0.729	0.885	0.801	0.883	0.977
2	1	500	0.629	0.695	0.834	0.731	0.794	0.916	0.878	0.930	0.985
2	1	1000	0.724	0.765	0.874	0.817	0.866	0.941	0.945	0.964	0.991

Panel B: With Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol4	Sol5	Sol6	Sol4	Sol5	Sol6	Sol4	Sol5	Sol6
2	1	10	0.355	0.448	0.650	0.418	0.507	0.705	0.523	0.604	0.771
2	1	20	0.376	0.493	0.740	0.445	0.559	0.808	0.556	0.668	0.897
2	1	50	0.482	0.620	0.855	0.567	0.706	0.914	0.701	0.830	0.974
2	1	100	0.618	0.731	0.905	0.702	0.812	0.951	0.830	0.919	0.993
2	1	200	0.721	0.810	0.939	0.804	0.889	0.975	0.922	0.967	0.997
2	1	300	0.762	0.825	0.943	0.829	0.904	0.975	0.943	0.973	0.995
2	1	500	0.818	0.870	0.950	0.894	0.934	0.981	0.968	0.987	0.998
2	1	1000	0.853	0.891	0.954	0.923	0.951	0.985	0.985	0.991	0.998

Table B.5: Empirical Coverages for the Simple Mean Efficiency using FDH Estimator
(continued)

Panel A: Without Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol1	Sol2	Sol3	Sol1	Sol2	Sol3	Sol1	Sol2	Sol3
3	1	10	0.047	0.077	0.232	0.059	0.095	0.282	0.084	0.133	0.343
3	1	20	0.052	0.095	0.346	0.068	0.121	0.420	0.111	0.189	0.553
3	1	50	0.066	0.123	0.466	0.089	0.168	0.598	0.156	0.279	0.787
3	1	100	0.081	0.162	0.585	0.119	0.246	0.753	0.227	0.425	0.929
3	1	200	0.191	0.321	0.766	0.265	0.434	0.889	0.434	0.669	0.991
3	1	300	0.259	0.420	0.817	0.351	0.550	0.921	0.567	0.786	0.994
3	1	500	0.386	0.542	0.872	0.501	0.683	0.962	0.717	0.887	0.999
3	1	1000	0.557	0.702	0.921	0.683	0.822	0.979	0.866	0.958	1.000

Panel B: With Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol4	Sol5	Sol6	Sol4	Sol5	Sol6	Sol4	Sol5	Sol6
3	1	10	0.269	0.321	0.467	0.319	0.368	0.504	0.416	0.460	0.569
3	1	20	0.288	0.390	0.688	0.357	0.457	0.746	0.479	0.562	0.826
3	1	50	0.346	0.505	0.874	0.424	0.591	0.931	0.566	0.740	0.981
3	1	100	0.433	0.645	0.948	0.538	0.750	0.981	0.701	0.884	0.997
3	1	200	0.601	0.810	0.989	0.704	0.894	0.996	0.873	0.976	1.000
3	1	300	0.695	0.859	0.992	0.794	0.930	0.999	0.911	0.987	1.000
3	1	500	0.784	0.921	0.995	0.870	0.972	1.000	0.969	0.997	1.000
3	1	1000	0.868	0.950	0.995	0.933	0.979	1.000	0.982	1.000	1.000

Table B.5: Empirical Coverages for the Simple Mean Efficiency using FDH Estimator
(continued)

Panel A: Without Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol1	Sol2	Sol3	Sol1	Sol2	Sol3	Sol1	Sol2	Sol3
4	1	10	0.022	0.036	0.113	0.026	0.046	0.130	0.036	0.065	0.159
4	1	20	0.013	0.024	0.153	0.017	0.029	0.200	0.025	0.049	0.275
4	1	50	0.013	0.021	0.276	0.015	0.035	0.381	0.027	0.069	0.575
4	1	100	0.009	0.023	0.333	0.013	0.037	0.483	0.029	0.088	0.764
4	1	200	0.011	0.035	0.454	0.018	0.054	0.673	0.044	0.142	0.929
4	1	300	0.024	0.062	0.598	0.036	0.113	0.824	0.090	0.267	0.985
4	1	500	0.042	0.108	0.684	0.070	0.172	0.891	0.147	0.396	0.998
4	1	1000	0.123	0.250	0.824	0.185	0.387	0.965	0.347	0.686	1.000

Panel B: With Data Sharpening

p	q	n	0.90			0.95			0.99		
			Sol4	Sol5	Sol6	Sol4	Sol5	Sol6	Sol4	Sol5	Sol6
4	1	10	0.216	0.240	0.298	0.289	0.312	0.358	0.447	0.462	0.497
4	1	20	0.163	0.207	0.457	0.210	0.259	0.499	0.324	0.379	0.577
4	1	50	0.177	0.269	0.741	0.230	0.339	0.821	0.368	0.486	0.911
4	1	100	0.171	0.306	0.857	0.226	0.405	0.939	0.371	0.590	0.984
4	1	200	0.229	0.436	0.957	0.301	0.566	0.989	0.478	0.788	1.000
4	1	300	0.319	0.572	0.986	0.418	0.713	0.999	0.614	0.907	1.000
4	1	500	0.399	0.687	0.997	0.505	0.829	1.000	0.714	0.963	1.000
4	1	1000	0.600	0.859	1.000	0.722	0.947	1.000	0.890	0.997	1.000

Table B.5: Empirical Coverages for the Simple Mean Efficiency using FDH Estimator
(continued)

Panel A: Without Data Sharpening													
p	q	n	0.90			0.95			0.99				
			Sol1	Sol2	Sol3	Sol1	Sol2	Sol3	Sol1	Sol2	Sol3		
5	1	50	0.004	0.009	0.187	0.005	0.011	0.251	0.009	0.025	0.381		
5	1	100	0.002	0.003	0.212	0.002	0.005	0.364	0.004	0.017	0.607		
5	1	200	0.003	0.008	0.347	0.004	0.017	0.576	0.011	0.037	0.865		
5	1	300	0.004	0.013	0.418	0.006	0.020	0.686	0.014	0.055	0.951		
5	1	500	0.009	0.025	0.603	0.016	0.050	0.874	0.034	0.129	0.998		
5	1	1000	0.025	0.072	0.839	0.038	0.136	0.982	0.090	0.364	1.000		

Panel B: With Data Sharpening													
p	q	n	0.90			0.95			0.99				
			Sol4	Sol5	Sol6	Sol4	Sol5	Sol6	Sol4	Sol5	Sol6		
5	1	50	0.120	0.165	0.606	0.164	0.229	0.669	0.303	0.385	0.772		
5	1	100	0.105	0.173	0.807	0.148	0.248	0.888	0.270	0.428	0.962		
5	1	200	0.124	0.264	0.938	0.180	0.380	0.981	0.341	0.608	0.998		
5	1	300	0.150	0.325	0.980	0.217	0.460	0.998	0.386	0.720	1.000		
5	1	500	0.228	0.493	0.997	0.307	0.663	1.000	0.521	0.890	1.000		
5	1	1000	0.397	0.767	1.000	0.522	0.887	1.000	0.751	0.985	1.000		

Table B.5: Empirical Coverages for the Simple Mean Efficiency using FDH Estimator
(continued)

Panel A: Without Data Sharpening											
p	q	n	0.90			0.95			0.99		
			Sol1	Sol2	Sol3	Sol1	Sol2	Sol3	Sol1	Sol2	Sol3
7	1	50	0.003	0.004	0.139	0.003	0.004	0.177	0.003	0.011	0.251
7	1	100	0.001	0.002	0.149	0.001	0.003	0.234	0.001	0.005	0.370
7	1	200	0.001	0.004	0.272	0.002	0.005	0.447	0.004	0.011	0.700
7	1	300	0.001	0.002	0.264	0.001	0.003	0.475	0.002	0.006	0.800
7	1	500	0.002	0.003	0.498	0.003	0.004	0.797	0.003	0.017	0.984
7	1	1000	0.007	0.017	0.785	0.009	0.031	0.981	0.018	0.062	1.000
Panel B: With Data Sharpening											
p	q	n	0.90			0.95			0.99		
			Sol4	Sol5	Sol6	Sol4	Sol5	Sol6	Sol4	Sol5	Sol6
7	1	50	0.174	0.200	0.472	0.256	0.287	0.539	0.445	0.476	0.661
7	1	100	0.088	0.124	0.647	0.139	0.184	0.729	0.281	0.348	0.823
7	1	200	0.133	0.206	0.898	0.201	0.298	0.953	0.369	0.517	0.987
7	1	300	0.104	0.186	0.947	0.161	0.287	0.982	0.313	0.497	0.998
7	1	500	0.163	0.308	0.997	0.233	0.451	1.000	0.423	0.713	1.000
7	1	1000	0.239	0.525	1.000	0.333	0.700	1.000	0.565	0.915	1.000

Table B.6: Simple Mean Efficiency and Bias using FDH Estimator

Panel A: Without Data Sharpening

p	q	n	$\hat{\mu}_n$	$\hat{B}_n$	$\hat{\mu}_n - \hat{B}_n$	μ_s	μ
1	1	10	1.2474	-0.2312	1.4787	1.8009	1.7979
1	1	20	1.3494	-0.2402	1.5896	1.7994	1.7979
1	1	50	1.4675	-0.2152	1.6828	1.7966	1.7979
1	1	100	1.5447	-0.1856	1.7302	1.7956	1.7979
1	1	200	1.6084	-0.1518	1.7602	1.7967	1.7979
1	1	300	1.6395	-0.1320	1.7716	1.7982	1.7979
1	1	500	1.6708	-0.1085	1.7793	1.7972	1.7979
1	1	1000	1.7058	-0.0823	1.7881	1.7979	1.7979

Panel B: With Data Sharpening

p	q	n	$\hat{\mu}_n$	$\hat{B}_n$	$\hat{\mu}_n - \hat{B}_n$	μ_s	μ
1	1	10	1.4215	-0.2439	1.6654	1.8009	1.7979
1	1	20	1.4556	-0.2467	1.7023	1.7994	1.7979
1	1	50	1.5198	-0.2176	1.7375	1.7966	1.7979
1	1	100	1.5740	-0.1866	1.7607	1.7956	1.7979
1	1	200	1.6247	-0.1523	1.7770	1.7967	1.7979
1	1	300	1.6511	-0.1323	1.7834	1.7982	1.7979
1	1	500	1.6783	-0.1087	1.7870	1.7972	1.7979
1	1	1000	1.7100	-0.0824	1.7924	1.7979	1.7979

NOTE: $\hat{\mu}_n$: Mean efficiency constructed using n observations; $\hat{B}_n$: Bias constructed using n observations; μ_s : Sample mean of true efficiency; μ : Population mean of true efficiency.

Table B.6: Simple Mean Efficiency and Bias using FDH Estimator (continued)

Panel A: Without Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
2	1	10	1.0936	-0.1649	1.2584	1.7933	1.7979
2	1	20	1.1459	-0.2131	1.3590	1.7986	1.7979
2	1	50	1.2380	-0.2662	1.5042	1.7975	1.7979
2	1	100	1.3107	-0.2826	1.5933	1.7992	1.7979
2	1	200	1.3867	-0.2773	1.6640	1.7986	1.7979
2	1	300	1.4225	-0.2657	1.6881	1.7973	1.7979
2	1	500	1.4703	-0.2487	1.7190	1.7980	1.7979
2	1	1000	1.5291	-0.2205	1.7496	1.7988	1.7979

Panel B: With Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
2	1	10	1.3719	-0.1859	1.5578	1.7933	1.7979
2	1	20	1.3638	-0.2319	1.5957	1.7986	1.7979
2	1	50	1.3886	-0.2789	1.6675	1.7975	1.7979
2	1	100	1.4185	-0.2910	1.7095	1.7992	1.7979
2	1	200	1.4609	-0.2824	1.7433	1.7986	1.7979
2	1	300	1.4826	-0.2695	1.7521	1.7973	1.7979
2	1	500	1.5152	-0.2513	1.7665	1.7980	1.7979
2	1	1000	1.5592	-0.2221	1.7813	1.7988	1.7979

NOTE: $\hat{\mu}_{n_\kappa}$: Mean efficiency constructed using n_κ observations; $\hat{B}_n$: Bias constructed using n observations; μ_s : Sample mean of true efficiency; μ : Population mean of true efficiency.

Table B.6: Simple Mean Efficiency and Bias using FDH Estimator (continued)

Panel A: Without Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
3	1	10	1.0357	-0.0987	1.1344	1.8014	1.7979
3	1	20	1.0644	-0.1497	1.2141	1.7985	1.7979
3	1	50	1.1198	-0.2268	1.3466	1.7956	1.7979
3	1	100	1.1677	-0.2707	1.4384	1.7967	1.7979
3	1	200	1.2314	-0.3057	1.5370	1.7994	1.7979
3	1	300	1.2633	-0.3153	1.5786	1.7978	1.7979
3	1	500	1.3092	-0.3206	1.6298	1.7977	1.7979
3	1	1000	1.3685	-0.3148	1.6833	1.7980	1.7979

Panel B: With Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
3	1	10	1.3622	-0.1184	1.4806	1.8014	1.7979
3	1	20	1.3485	-0.1711	1.5196	1.7985	1.7979
3	1	50	1.3460	-0.2488	1.5948	1.7956	1.7979
3	1	100	1.3545	-0.2899	1.6444	1.7967	1.7979
3	1	200	1.3784	-0.3209	1.6992	1.7994	1.7979
3	1	300	1.3916	-0.3282	1.7198	1.7978	1.7979
3	1	500	1.4146	-0.3308	1.7454	1.7977	1.7979
3	1	1000	1.4483	-0.3219	1.7702	1.7980	1.7979

NOTE: $\hat{\mu}_{n_\kappa}$: Mean efficiency constructed using n_κ observations; $\hat{B}_n$: Bias constructed using n observations; μ_s : Sample mean of true efficiency; μ : Population mean of true efficiency.

Table B.6: Simple Mean Efficiency and Bias using FDH Estimator (continued)

Panel A: Without Data Sharpening

p	q	n	$\hat{\mu}_n$	$\hat{B}_n$	$\hat{\mu}_n - \hat{B}_n$	μ_s	μ
4	1	10	1.0109	-0.0435	1.0544	1.8012	1.7979
4	1	20	1.0220	-0.0697	1.0917	1.7958	1.7979
4	1	50	1.0484	-0.1317	1.1801	1.7974	1.7979
4	1	100	1.0755	-0.1837	1.2592	1.7967	1.7979
4	1	200	1.1095	-0.2342	1.3437	1.7973	1.7979
4	1	300	1.1366	-0.2633	1.3999	1.7982	1.7979
4	1	500	1.1677	-0.2943	1.4619	1.7979	1.7979
4	1	1000	1.2135	-0.3226	1.5360	1.7979	1.7979

Panel B: With Data Sharpening

p	q	n	$\hat{\mu}_n$	$\hat{B}_n$	$\hat{\mu}_n - \hat{B}_n$	μ_s	μ
4	1	10	1.3675	-0.0540	1.4215	1.8012	1.7979
4	1	20	1.3432	-0.0843	1.4275	1.7958	1.7979
4	1	50	1.3234	-0.1517	1.4752	1.7974	1.7979
4	1	100	1.3195	-0.2058	1.5253	1.7967	1.7979
4	1	200	1.3235	-0.2561	1.5796	1.7973	1.7979
4	1	300	1.3308	-0.2844	1.6152	1.7982	1.7979
4	1	500	1.3386	-0.3136	1.6522	1.7979	1.7979
4	1	1000	1.3572	-0.3387	1.6960	1.7979	1.7979

NOTE: $\hat{\mu}_{n_\kappa}$: Mean efficiency constructed using n_κ observations; $\hat{B}_n$: Bias constructed using n observations; μ_s : Sample mean of true efficiency; μ : Population mean of true efficiency.

Table B.6: Simple Mean Efficiency and Bias using FDH Estimator (continued)

Panel A: Without Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
5	1	50	1.0188	-0.0782	1.0971	1.7965	1.7979
5	1	100	1.0365	-0.1208	1.1573	1.7959	1.7979
5	1	200	1.0535	-0.1737	1.2272	1.7966	1.7979
5	1	300	1.0741	-0.2034	1.2775	1.7980	1.7979
5	1	500	1.0959	-0.2430	1.3389	1.7981	1.7979
5	1	1000	1.1360	-0.2913	1.4273	1.7981	1.7979

Panel B: With Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
5	1	50	1.3286	-0.0934	1.4220	1.7965	1.7979
5	1	100	1.3171	-0.1407	1.4578	1.7959	1.7979
5	1	200	1.3094	-0.1971	1.5065	1.7966	1.7979
5	1	300	1.3116	-0.2277	1.5393	1.7980	1.7979
5	1	500	1.3154	-0.2678	1.5832	1.7981	1.7979
5	1	1000	1.3277	-0.3153	1.6430	1.7981	1.7979

NOTE: $\hat{\mu}_{n_\kappa}$: Mean efficiency constructed using n_κ observations; $\hat{B}_n$: Bias constructed using n observations; μ_s : Sample mean of true efficiency; μ : Population mean of true efficiency.

Table B.6: Simple Mean Efficiency and Bias using FDH Estimator (continued)

Panel A: Without Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
7	1	50	1.0082	-0.0393	1.0475	1.8019	1.7979
7	1	100	1.0126	-0.0648	1.0774	1.7979	1.7979
7	1	200	1.0230	-0.1003	1.1233	1.7985	1.7979
7	1	300	1.0313	-0.1261	1.1574	1.7978	1.7979
7	1	500	1.0425	-0.1623	1.2048	1.7982	1.7979
7	1	1000	1.0645	-0.2169	1.2814	1.7985	1.7979

Panel B: With Data Sharpening

p	q	n	$\hat{\mu}_{n_\kappa}$	$\hat{B}_n$	$\hat{\mu}_{n_\kappa} - \hat{B}_n$	μ_s	μ
7	1	50	1.3523	-0.0490	1.4013	1.8019	1.7979
7	1	100	1.3384	-0.0791	1.4175	1.7979	1.7979
7	1	200	1.3342	-0.1202	1.4544	1.7985	1.7979
7	1	300	1.3269	-0.1492	1.4761	1.7978	1.7979
7	1	500	1.3227	-0.1892	1.5119	1.7982	1.7979
7	1	1000	1.3214	-0.2477	1.5691	1.7985	1.7979

NOTE: $\hat{\mu}_{n_\kappa}$: Mean efficiency constructed using n_κ observations; $\hat{B}_n$: Bias constructed using n observation; μ_s : Sample mean of true efficiency; μ : Population mean of true efficiency.

Table B.7: Mean Standard Deviation of Estimated Efficiency using FDH Estimator

p	q	n	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	σ_s	σ
1	1	10	0.3574	0.4318	0.6712	0.3165	0.4068	0.6269	0.5813	0.6028
1	1	20	0.4383	0.5039	0.7025	0.3836	0.4608	0.6440	0.5976	0.6028
1	1	50	0.4936	0.5404	0.6778	0.4541	0.5056	0.6355	0.5970	0.6028
1	1	100	0.5248	0.5575	0.6574	0.4988	0.5335	0.6295	0.5993	0.6028
1	1	200	0.5473	0.5683	0.6405	0.5311	0.5529	0.6232	0.6014	0.6028
1	1	300	0.5572	0.5729	0.6322	0.5451	0.5612	0.6194	0.6026	0.6028
1	1	500	0.5655	0.5760	0.6222	0.5573	0.5679	0.6134	0.6022	0.6028
1	1	1000	0.5757	0.5816	0.6144	0.5708	0.5768	0.6093	0.6025	0.6028

NOTE: σ_s : Sample standard deviation of true efficiency; σ : Population standard deviation of true efficiency.

Table B.7: Mean Standard Deviation of Estimated Efficiency using FDH Estimator (continued)

p	q	n	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	σ_s	σ
2	1	10	0.1925	0.2574	0.5335	0.2461	0.3202	0.5895	0.5806	0.6028
2	1	20	0.2777	0.3539	0.6760	0.2693	0.3603	0.6662	0.5928	0.6028
2	1	50	0.3659	0.4547	0.7480	0.3191	0.4263	0.6990	0.5971	0.6028
2	1	100	0.4189	0.5066	0.7587	0.3675	0.4702	0.7047	0.6007	0.6028
2	1	200	0.4582	0.5363	0.7415	0.4120	0.5003	0.6925	0.6022	0.6028
2	1	300	0.4760	0.5456	0.7240	0.4348	0.5120	0.6800	0.6017	0.6028
2	1	500	0.4964	0.5555	0.7057	0.4618	0.5261	0.6686	0.6026	0.6028
2	1	1000	0.5180	0.5631	0.6809	0.4918	0.5397	0.6527	0.6029	0.6028

NOTE: σ_s : Sample standard deviation of true efficiency; σ : Population standard deviation of true efficiency.

Table B.7: Mean Standard Deviation of Estimated Efficiency using FDH Estimator (continued)

p	q	n	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	σ_s	σ
3	1	10	0.0969	0.1401	0.3518	0.2275	0.2730	0.4950	0.5802	0.6028
3	1	20	0.1709	0.2302	0.5598	0.2329	0.2973	0.6216	0.5928	0.6028
3	1	50	0.2596	0.3471	0.7336	0.2571	0.3607	0.7330	0.5965	0.6028
3	1	100	0.3114	0.4140	0.7879	0.2811	0.4052	0.7592	0.5985	0.6028
3	1	200	0.3629	0.4753	0.8131	0.3168	0.4518	0.7678	0.6016	0.6028
3	1	300	0.3873	0.4999	0.8121	0.3374	0.4712	0.7624	0.6019	0.6028
3	1	500	0.4159	0.5254	0.8037	0.3653	0.4932	0.7525	0.6030	0.6028
3	1	1000	0.4479	0.5476	0.7806	0.4007	0.5142	0.7319	0.6030	0.6028

NOTE: σ_s : Sample standard deviation of true efficiency; σ : Population standard deviation of true efficiency.

Table B.7: Mean Standard Deviation of Estimated Efficiency using FDH Estimator (continued)

p	q	n	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	σ_s	σ
4	1	10	0.0356	0.0568	0.1606	0.2162	0.2372	0.3473	0.5814	0.6028
4	1	20	0.0759	0.1044	0.3175	0.2106	0.2369	0.4529	0.5865	0.6028
4	1	50	0.1530	0.2039	0.5704	0.2159	0.2683	0.6312	0.5972	0.6028
4	1	100	0.2031	0.2752	0.6920	0.2276	0.3087	0.7184	0.5996	0.6028
4	1	200	0.2483	0.3421	0.7656	0.2429	0.3539	0.7645	0.6002	0.6028
4	1	300	0.2767	0.3825	0.8004	0.2569	0.3838	0.7854	0.6021	0.6028
4	1	500	0.3082	0.4265	0.8274	0.2753	0.4176	0.7990	0.6016	0.6028
4	1	1000	0.3483	0.4748	0.8370	0.3044	0.4556	0.7967	0.6031	0.6028

NOTE: σ_s : Sample standard deviation of true efficiency; σ : Population standard deviation of true efficiency.

Table B.7: Mean Standard Deviation of Estimated Efficiency using FDH Estimator (continued)

p	q	n	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	σ_s	σ
5	1	50	0.0922	0.1224	0.4230	0.2053	0.2316	0.5274	0.5993	0.6028
5	1	100	0.1372	0.1841	0.5817	0.2088	0.2544	0.6511	0.6005	0.6028
5	1	200	0.1792	0.2503	0.7029	0.2167	0.2941	0.7454	0.6009	0.6028
5	1	300	0.2029	0.2879	0.7539	0.2234	0.3198	0.7815	0.6013	0.6028
5	1	500	0.2335	0.3374	0.8082	0.2347	0.3565	0.8181	0.6025	0.6028
5	1	1000	0.2745	0.4005	0.8581	0.2546	0.4054	0.8478	0.6029	0.6028

NOTE: σ_s : Sample standard deviation of true efficiency; σ : Population standard deviation of true efficiency.

Table B.7: Mean Standard Deviation of Estimated Efficiency using FDH Estimator (continued)

p	q	n	Sol1	Sol2	Sol3	Sol4	Sol5	Sol6	σ_s	σ
7	1	50	0.0425	0.0586	0.2621	0.2107	0.2220	0.4221	0.6013	0.6028
7	1	100	0.0746	0.1000	0.4305	0.2063	0.2253	0.5493	0.6008	0.6028
7	1	200	0.1064	0.1472	0.5807	0.2062	0.2411	0.6806	0.6016	0.6028
7	1	300	0.1254	0.1785	0.6563	0.2081	0.2576	0.7457	0.6016	0.6028
7	1	500	0.1506	0.2218	0.7411	0.2124	0.2852	0.8175	0.6023	0.6028
7	1	1000	0.1862	0.2861	0.8404	0.2218	0.3328	0.8968	0.6034	0.6028

NOTE: σ_s : Sample standard deviation of true efficiency; σ : Population standard deviation of true efficiency.

Appendix C Details and Additional Results from Simulations for CRS-DEA Estimator

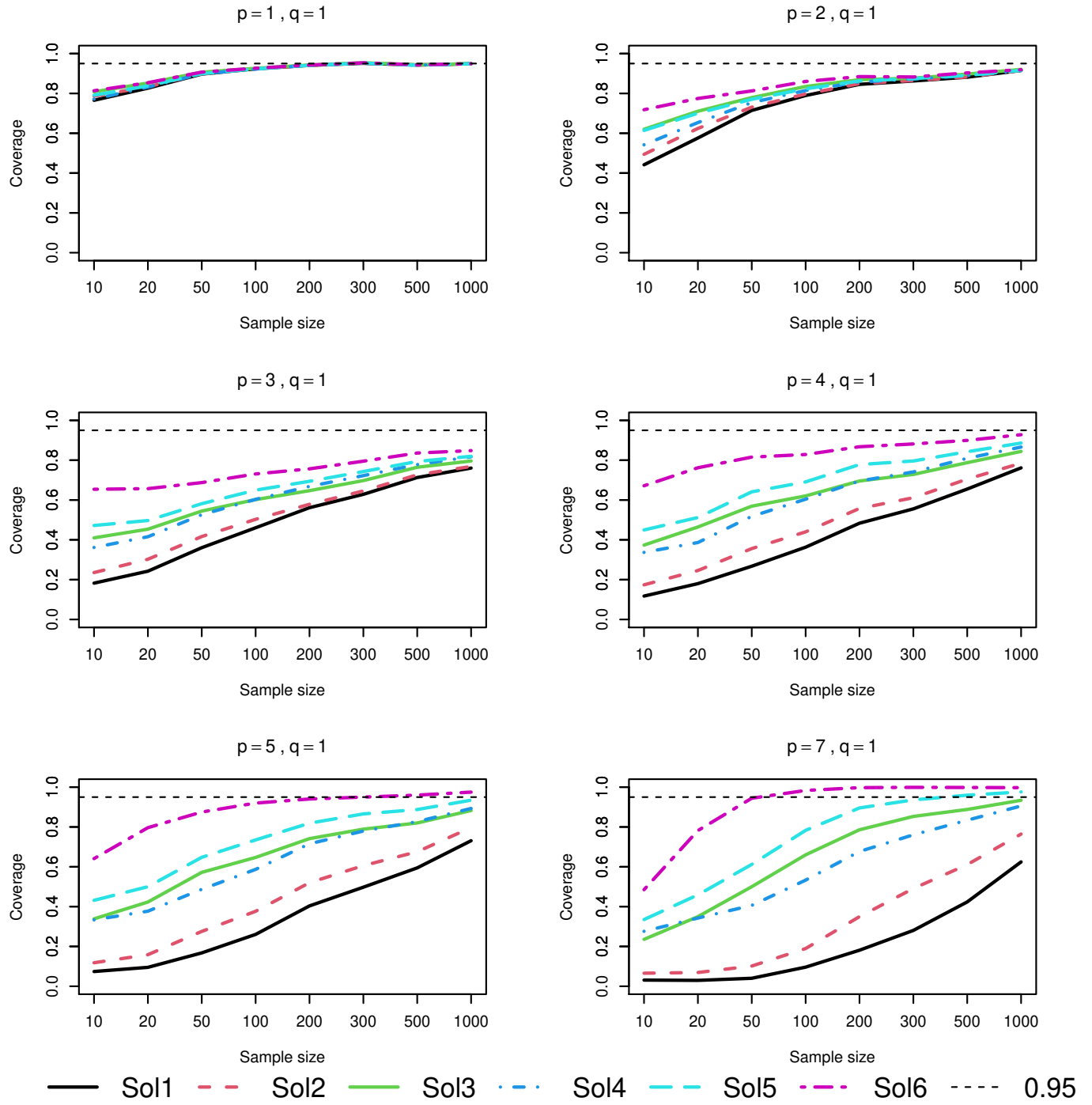
In this appendix, we conduct extensive Monte-Carlo experiments to illustrate the performance of our proposed method for CRS-DEA estimator. For CRS estimator, the technology needs to be globally constant returns to scale. Following [Kneip et al. \(2015\)](#), we use the following technology,

$$y_i^\partial(x_i) = \prod_{j=1}^p x_{ji}^{1/p}, \quad (\text{C.1})$$

The inefficiency is generated similarly as for VRS-DEA estimator presented in [Appendix A](#). Instead of using VRS-DEA estimator, here we use CRS-DEA estimator.

The results are summarized in [Figure C.1](#), confirming the improved performance of our proposed method for CRS-DEA estimator.

Figure C.1: Empirical Coverages for the Simple Mean Efficiency using CRS-DEA Estimator



References

- Kneip, A., L. Simar, and P. W. Wilson (2015), When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores, *Econometric Theory* 31, 394–422.
- Nguyen, H., L. Simar, and V. Zelenyuk (2022), Data sharpening for improving CLT approximations for DEA-type efficiency estimators, *European Journal of Operational Research*, forthcoming.
- Simar, L. and V. Zelenyuk (2020), Improving finite sample approximation by central limit theorems for estimates from Data Envelopment Analysis, *European Journal of Operational Research* 284, 1002–1015.