Interfaces with Other Disciplines

One-stage and two-stage DEA estimation of the effects of contextual variables

Andrew L. Johnson, Timo Kuosmanen

Abstract

Two-stage data envelopment analysis (2-DEA) is commonly used in productive efficiency analysis to estimate the effects of operational conditions and practices on performance. In this method the DEA efficiency estimates are regressed on contextual variables representing the operational conditions. We re-examine the statistical properties of the 2-DEA estimator, and find that it is statistically consistent under more general conditions than earlier studies assume. We further show that the finite sample bias of DEA in the first stage carries over to the second stage regression, causing bias in the estimated coefficients of the contextual variables. This bias is particularly severe when the contextual variables are correlated with inputs. To address this shortcoming, we apply the result that DEA can be formulated as a constrained special case of the convex nonparametric least squares (CNLS) regression. Applying the CNLS formulation, we develop a new semi-nonparametric one-stage estimator for the coefficients of the contextual variables that directly incorporates contextual variables to the standard DEA problem. The proposed method is hence referred to as one-stage DEA (1-DEA). Evidence from Monte Carlo simulations suggests that the new 1-DEA estimator performs systematically better than the conventional 2-DEA estimator both in deterministic and noisy scenarios.

Article history:
Received 22 March 2011
Accepted 12 January 2012
Available online 24 January 2012

Keywords:
Data envelopment analysis
Partial linear model
Semi-nonparametric regression

1. Introduction

Technical efficiency requires a firm to produce the maximum output for a given level of input use. Firm's ability to operate efficiently often depends on operational conditions and practices, such as the external operational environment in which production occurs, the internal characteristics of the firms such as the type and vintage of technology, and the managerial practices. Using the terminology of Banker and Natarajan (2008), we will henceforth refer to variables that characterize operational conditions and practices as contextual variables.

Estimating the effects of contextual variables on efficiency can provide valuable insight to managers who develop business strategies or make decisions on operational practices, and for policy makers who may influence the external operating environment of firms through standards, regulations, taxes, subsidies, and other policy measures. This topic has attracted deserved attention in the literature of data envelopment analysis (DEA). Ray (1988) was the first to apply the two-stage DEA method (henceforth 2-DEA) where the efficient frontier and the firm-level efficiency scores are estimated with DEA in the first stage, and the estimated efficiency scores are regressed on contextual variables in the second stage. A variety of regression techniques have been used, including the classic ordinary least squares (OLS) and the maximum likelihood (ML) based probit, logit, and truncated regression (e.g., Simar and Wilson, 2007).

Although the 2-DEA method has proved useful in a large number of applications, its statistical foundation is currently subject to intensive debate (e.g., Hoff, 2007; McDonald, 2009; Simar and Wilson, 2007). Most notably, Simar and Wilson (2007) have sharply criticized the 2-DEA method for a lack of a coherent data generating process (DGP) and for the bias and serial correlation of the DEA efficiency estimates. Most importantly, they argue that the conventional methods of statistical inference are invalid in the second stage regression. To address these problems, the authors propose the use of a bootstrap method to correct for the small sample bias and serial correlation of the DEA efficiency estimates. Further, they advocate the use of the truncated regression model that takes into account explicitly the bounded domain of the DEA efficiency estimates.

Another notable examination of the statistical basis of the 2-DEA method is Banker and Natarajan (2008). They show that the 2-DEA estimator for the contextual variables is statistically consistent under certain assumptions and regularity conditions. In contrast to Simar and Wilson who do not consider stochastic noise, Banker and Natarajan introduce a noise term that has a truncated...
distribution, following the DEA+ approach by Gstach (1998). Further, Banker and Natarajan assume the sign of the contextual variables is known beforehand, and that the vector of contextual variables is independently distributed relative to the input vector. They also report Monte Carlo simulations, benchmarking 2-DEA against both one-stage and two-stage parametric methods. The results indicate that if the exact functional form of the underlying DGP is known, the one-stage parametric estimator outperforms the 2-DEA method. However, if the functional form is misspecified, the 2-DEA method outperforms the parametric estimators. The simulation results also show that the two-stage method performs reasonably well when the contextual variables are uncorrelated with the inputs. However, as the correlation between the inputs and contextual variables increases, the performance of the 2-DEA estimator deteriorates.

While the recent studies shed new light on the performance of the 2-DEA method under certain assumptions, the sharp contrast in the conclusions of Simar and Wilson (2007) and Banker and Natarajan (2008) calls for further examination of 2-DEA under more general assumptions about the DGP. In particular, the effects of finite sample bias, stochastic noise, and the correlation between the inputs and contextual variables deserve further elaboration.

The contribution of the present paper is threefold, and can be summarized as follows: (1) we further elaborate the statistical properties of the 2-DEA estimator under more general assumptions, (2) we develop a new one-stage DEA (1-DEA) method that facilitates joint estimation of the frontier and the effects of contextual variables, and (3) we compare the performance of the conventional 2-DEA method and the proposed 1-DEA method in the controlled environment of Monte Carlo simulations.

Regarding the first contribution, we show that the OLS regression of the DEA efficiency scores on the contextual variables provides a statistically consistent estimator of the coefficients of contextual variables under more general assumptions than those imposed by Banker and Natarajan (2008). In particular, we show that consistency of the 2-DEA estimator does not require the contextual variables be uncorrelated with inputs. Unlike the regression methods such as stochastic frontier analysis (SFA), the DEA efficiency estimator is not subject to the omitted variable bias in the first stage if the effect of the contextual variables has a finite maximum and the sample size is sufficiently large. However, the small sample bias of the DEA estimator will carry over to the second stage regression. We derive a formal expression for the bias of the second stage estimator, which reveals the link between the bias and the correlation between the input and contextual variables. Our results explain why the precision of 2-DEA estimator deteriorates when the correlation of the inputs and the contextual variables increases, as reported in Banker and Natarajan’s Monte Carlo simulation results.

Regarding the second contribution, the results for 2-DEA and the parallel findings from the SFA literature (e.g., Wang and Schmidt, 2002) suggest that unbiased and efficient estimation of the effects of contextual variables requires joint estimation of the frontier and the coefficients of the contextual variables by using a one-stage method. To our knowledge, such a one-stage semi-nonparametric DEA-style estimator has not been considered before. To develop such an estimator, we apply insights from our earlier study (Kuosmanen and Johnson, 2010), where we showed that standard DEA has a regression interpretation as a constrained variant of the convex nonparametric least squares (CNLS) regression (Hildreth, 1954; Hanson and Pledger, 1976; Kuosmanen, 2008).

Utilizing this regression interpretation of DEA, we show that the regression model for the contextual variables can be integrated into the standard DEA formulation to develop a new one-stage semi-nonparametric estimator. The main advantage of 1-DEA is that it takes into account the correlation of inputs and contextual variables in the simultaneous estimation of the frontier and the effects of contextual variables.

CNLS regression was first used in productive efficiency analysis as the first stage of the StoNED method (stochastic nonparametric envelopment of data) by Kuosmanen (2006) and Kuosmanen and Kortelainen (in press), which combines the axiomatic DEA-style frontier with the stochastic SFA-style treatment of inefficiency and noise. Parallel to this paper, in Johnson and Kuosmanen (2011) we introduce the contextual variables to the StoNED model and develop a method referred to as stochastic semi-nonparametric envelopment of 2-variables data (StoNEZD). The StoNEZD method is similar to the 1-DEA introduced in this paper in that it jointly estimates the frontier and the contextual variables using CNLS regression. The key difference is the truncated noise term (Gstach, 1998) assumed in this paper. One possible source of truncation is the use of outlier detection methods where observations that lie too far above the estimated CNLS curve are classified as outliers and removed from the sample. Further, the classic deterministic model that excludes the noise term is a special case of the truncated model where the truncation point is set to zero. Although the StoNEZD estimator introduced in Johnson and Kuosmanen (2011) is consistent even if the noise term is truncated, in this paper we show how the truncation of the noise term can be explicitly taken into account in the estimation.

Regarding the third contribution, we complement the asymptotic properties established in this paper by examining the finite sample performance of the 2-DEA and 1-DEA estimators via Monte Carlo simulation. To ensure the DGP used in the simulations is fair and comparable with earlier published results, we replicate the DGP considered by Banker and Natarajan (2008), also examined in Johnson and Kuosmanen (2011). In a wide variety of scenarios, the results show that the proposed 1-DEA estimator yields more precise estimates than 2-DEA even if the truncation point is wrongly specified. Further, our results suggest that taking into account the truncation of the noise term can improve the finite sample performance of the 1-DEA estimator, especially in small samples and when the probability of truncation is high. Our results provide strong support for using the proposed one-stage methods for joint estimation of the frontier and the effects of the contextual variables.

Taking the heterogeneity of firms and their operating environments into account is important in virtually all thinkable empirical applications of productive efficiency analysis. If the heterogeneity is ignored, firms operating under favorable conditions appear more efficient than firms operating in a harsh environment. One notable application of the one-stage semi-nonparametric approach considered in this paper concerns the government regulation of electricity distribution firms in Finland, documented in Kuosmanen et al. (2010) and Kuosmanen (2011). To avoid the abuse of local monopoly power in electricity distribution, the government regulators in several countries use DEA or SFA methods for estimating the efficient cost level and for setting efficiency improvement targets. In Finland, the regulator has decided to adopt the StoNED method by Kuosmanen and Kortelainen (in press), based on the recommendations by Kuosmanen et al. (2010). One of the challenges faced by the Finnish regulator concerns the heterogeneity of firms operating in urban, sub-urban, or rural areas. To better account for this heterogeneity in the efficiency analysis, Kuosmanen et al. (2010) introduce the proportion of underground cables in the intermediate voltage (1–70 kV) network as a contextual variable. The effect
of this variable is estimated using the one-stage approach developed in this study.

The rest of the paper is organized as follows. Section 2 introduces the theoretical model to be estimated, together with the necessary notation and assumptions. Section 3 discusses the conventional estimation procedures for handling production contexts and proves consistency of the 2-DEA estimator for quantifying the effect of contextual variables under more general conditions than previously literature. Section 4 develops the one-stage 1-DEA approach. Section 5 describes the Monte Carlo simulation results to demonstrate the performance of the proposed one-stage method compared to direct OLS and 2-DEA. Section 6 presents concluding remarks and offers suggestions for future research. The formal proofs of all theorems are provided in Appendix 1 and some additional Monte Carlo evidence is reported for the interval or ratio scale. Matrix \( B \) has full column rank. However, we do allow any two columns of \( B \) to be correlated. Unlike Banker and Natarajan (2008), we do not assume statistical independence of inputs and contextual variables.

For the parametric part of the model (i.e. \( Z \)) we assume the linear functional form, noting that the elements of \( Z \) can first be transformed by using a suitable data transformation (e.g., exponential or logarithmic). We can model nonlinear effects of contextual variables by introducing quadratic, cubic or higher-order polynomial transformations of the original contextual variables as additional columns of \( Z \). In addition to quantitative variables measured in the interval or ratio scale, matrix \( Z \) can contain categorical or ordinal information (for example, groups of firms), coded by using binary numbers. In econometrics, such binary variables are referred to as dummy variables, see Greene (2008) for a further discussion. To establish consistency, however, we must ensure that the elements of the domain \( D_z \) are bounded either from above or below such that the function \( \mathbf{z} \) has a finite maximum within \( D_z \). We further assume that the maximum value is non-negative: \( z \geq 0 \). A vector \( \mathbf{Z} \in \text{argmax}_{\mathbf{Z}} \mathbb{E} \mathbf{z} \) represents ideal operational conditions and practices (note that \( \mathbf{Z} \) may not be unique). Banker and Natarajan (2008) examine a special case where \( \mathbf{z} \leq 0 \) and \( \mathbf{Z} = 0 \). In contrast to their study, we do not restrict the signs of the elements of matrix \( Z \) or the parameter vector \( \mathbf{z} \) beforehand.

We do not impose any particular distributional assumptions on the random inefficiency and noise terms \( \mathbf{u} \) and \( \mathbf{v} \). Rather, we assume the existence of a well-behaved continuous joint density function \( f_{\mathbf{x},\mathbf{z},\mathbf{u},\mathbf{v}} \) with marginal densities \( f_{\mathbf{u}} \) and \( f_{\mathbf{v}} \), where \( f_{\mathbf{v}} \) is assumed to be symmetric with zero-mean, and \( f_{\mathbf{u}} \) is left-truncated at zero (i.e., \( f_{\mathbf{u}}(u) = 0 \) for all \( u < 0 \)). Random variables \( \mathbf{u} \) and \( \mathbf{v} \) are homoscedastic (i.e., have the same variance across observations), and denote the expected inefficiency as \( \mu = \mathbb{E}(u) > 0 \). Impacts of possible heteroskedasticity will be discussed below.

Following Gstach (1998) and Banker and Natarajan (2008), in this paper the noise term \( \mathbf{v} \) is assumed to be truncated at point \( \mathbf{v}^0 \) such that \( |\mathbf{v}| \leq \mathbf{v}^0 \mathbf{1} \), and \( f_{\mathbf{v}}(\mathbf{v}^0) > 0 \). We make this assumption for two reasons. The first one is methodological. Note that the classical deterministic model is obtained as a special case by setting \( \mathbf{v}^0 = 0 \). Further, statistical consistency of the DEA estimators that do not take noise explicitly into account critically depends on this assumption. It is obvious that the DEA estimator is inconsistent if the noise term is unbounded. The second reason is that truncation may result from the use of outlier detection methods in the prior data analysis (e.g., Rousseeuw and Leroy, 1987). If observations that lie too far above (or below) the estimated production frontier are classified as outliers and removed from the sample, the noise term \( \mathbf{v} \) is truncated from above (or below, see Johnson and McGinnis, 2008; Chen and Johnson, 2010). Knowing the criteria used in outlier detection, we can identify the truncation point \( \mathbf{v}^0 \). For notational convenience, we assume the truncation point \( \mathbf{v}^0 \) is constant across all observations. We are aware that the use of some outlier detection methods would imply the truncation points differ across firms. However, the proposed approach to modeling the truncation of the noise term naturally extends to the firm-specific truncation points \( \mathbf{v}^0_i \) (\( i = 1, \ldots, n \)), as discussed in Section 4.

We acknowledge that the truncated noise term is a peculiar assumption that does not conform to the usual statistical models of error. It is only valid in some special cases, such as the prior screening of outliers discussed above. For an analogous treatment

---

3 A cross-sectional model assumed throughout this paper uses data of \( n \) firms measured in the same time period. A data set that contains multiple measurements of each firm, observed over multiple time periods, is referred to as panel data.

4 See Simar and Wilson (2011, Appendix A) for critical discussion of this assumption.
of the standard normally distributed noise term, we refer to the parallel paper Johnson and Kuosmanen (2011).

To summarize, the model considered in this paper is based on Banker and Natarajan (2008), but we have relaxed two of their restrictive assumptions: (1) Inputs and the contextual variables are not assumed to be statistically independent. (2) The signs of the contextual variables or parameters δ are not restricted beforehand. In these respects it is evident that our theoretical model is more general than that of Banker and Natarajan (2008). Regarding (1), firms are usually aware of the operating conditions and practices, which will likely influence the firm’s decisions on the input use. Hence, we would usually expect columns of X and Z to be correlated. Regarding (2), it may be difficult to tell beforehand if a contextual variable has a positive or negative effect on performance, and estimating the sign has been a primary focus of many of the applications of 2-DEA (Simar and Wilson, 2007). Fortunately, we will show that these restrictive assumptions can be relaxed without sacrificing the statistical consistency of the 2-DEA estimator.

2.1. Direct OLS regression

To gain further insight regarding the correlation of inputs and contextual variables, we briefly examine a special case where the contextual variables Z, inputs X, inefficiency u and noise v are all statistically independent (i.e., none of these variables are correlated with another variable). In this case, if one is chiefly interested in estimating the marginal effects of the contextual variables δ, then a simple OLS regression can be used. Specifically, we may simply omit lnφ(x) from Eq. (1) and assign it to the disturbance term, together with the inefficiency and noise terms. We can apply a variable transformation lnφ(x) − u = a1 + w, where a = lnφ(1x) − μ is a constant and w = (lnφ(x) − u + v) − (lnφ(1x) − μ)1 is a vector of composite error terms that satisfies E(w) = 0. Thus, Eq. (1) can be rewritten as

\[ \ln y = a1 + Z\delta + w. \]  

(2)

Eq. (2) can be estimated by OLS, using the log-output as the dependent variable and the contextual variables Z as regressors.

Under the strong statistical independence assumption stated above, the direct OLS estimator δOLS of Eq. (2) is unbiased and consistent. Even though the composite error term w in Eq. (2) likely has a large variance, omitting lnφ(x) from the regression equation has no effect on the point estimate when the sample covariances of all pairs of columns take one column from X and one column from Z are equal to zero. Of course, statistical independence (i.e., zero population covariance) does not imply that the sample covariances of X and Z are equal to zero. Even if the sample correlation between the regressors Z and the inputs X is purely incidental, it will cause error in δOLS. In other words, we can expect the variance of the direct OLS estimator δOLS to be high. Taking the inputs X explicitly into account can help to decrease the variance considerably. We return to this point in more detail in the Monte Carlo simulations reported in Section 6.

Finally, we must stress that unbiasedness of the direct OLS estimator critically depends on the statistical independence of the inputs and contextual variables. Note that the disturbance term w is a function of inputs. Hence, if the assumption of statistical independence fails, then the composite disturbance term w is correlated with the regressors Z. In econometrics, this correlation is called endogeneity (see, e.g., Greene, 2008). To gain insight, consider the case of a single input x and a single contextual variable z, and assume the true coefficient δ is positive. If the correlation between x and z is positive, then the contextual variable z used as a regressor in (2) will capture not only its own effect, but also the effect of the input variable x. As a result, the coefficient δOLS is upward biased. Conversely, a negative correlation between x and z would result in a downward bias. The bias due to the correlation between x and z does not disappear as the sample size increases, and thus the direct OLS estimator is inconsistent.

3. Two-stage DEA

The literature on 2-DEA includes a number of variants (see, e.g., Simar and Wilson, 2007). This section follows the approach of Banker and Natarajan (2008).

The two stages of the 2-DEA method are the following. In the first stage, the frontier production function φ is estimated using the nonparametric DEA estimator formally stated as,

\[ \hat{\phi}_{\text{DEA}}(x) = \max_{\delta \in \mathbb{R}^m} \left\{ \sum_{h=1}^{n} \delta_h y_h \mid x \geq \sum_{h=1}^{n} x_h \delta_h = 1 \right\} \]  

(3)

The output error efficiency estimator of firm i is stated as \[ \hat{\eta}_i = \frac{y_i}{\hat{\phi}_{\text{DEA}}(x_i)}. \] Usually the efficiency estimator \[ \hat{\eta}_i \] is directly computed using

\[ \left( \hat{\eta}_i \right)^{-1} = \max_{\delta \in \mathbb{R}^m} \left\{ 0 \mid \delta \leq \sum_{h=1}^{n} x_h \delta_h \geq \sum_{h=1}^{n} y_h \delta_h = 1 \right\} \]  

(4)

In the second stage, the following linear regression equation is estimated using OLS or ML:

\[ \ln \hat{\eta}_i = \alpha + z_i \delta + \omega_i, \quad i = 1, \ldots, n, \]  

(5)

where the intercept α captures the expected inefficiency and the finite sample bias of the DEA estimator, and the composite disturbance term \[ \omega_i \] represents the noise term \[ \eta_i \] and the deviations of \[ \eta_i \] from the expected inefficiency \[ \mu \].

By construction, the DEA efficiency estimator \[ \hat{\eta}_i \] is bounded in the unit interval, and hence the dependent variable of Eq. (5) is truncated from above at \[ \ln \hat{\eta}_i = 0 \]. This is the reason why Simar and Wilson (2007) favor using truncated regression in the second stage (estimated by ML). However, the DGP they assume differs substantially from the model of Banker and Natarajan (2008) and its relaxed version described in Section 2. Specifically, truncation of the efficiency index is an essential assumption of the theoretical model of Simar and Wilson (2007). In contrast, Banker and Natarajan (2008) assume that the true efficiency index \[ \eta_i \] can be greater than one due to the noise term \[ \omega_i \]; it is possible that some observations are above the frontier when the output data are perturbed by positive noise. In this case, truncation of the dependent variable at \[ \ln \hat{\eta}_i = 0 \] is just an artifact of the DEA estimator and the deterministic assumption. Whereas Simar and Wilson (2007) find the standard 2-DEA estimator that overlooks the truncation to be biased and inconsistent, in the model of Banker and Natarajan (2008) the bias due to truncation vanishes asymptotically.

Banker and Natarajan (2008) correctly argue that the 2-DEA estimator is statistically consistent in the case of truncated noise. However, the assumptions required for consistency stated in Banker and Natarajan (2008) are overly restrictive. In fact, consistency of the 2-DEA estimator can be established under the relaxed set of assumptions introduced in Section 2 as follows (for clarity, we re-iterate the key assumptions). We state the result formally as follows:

**Theorem 1.** If the following five assumptions are satisfied:

(i) sequence \( \{y_i, x_i, z_i\}, i = 1, \ldots, n \) is a random sample of independent observations,

(ii) \( \lim_{n \to \infty} nZ \) is a positive definite matrix,

(iii) noise term v has a truncated distribution: \( |v| \leq \gamma^M, f_{\gamma}(\gamma^M) > 0 \),

(iv) elements of domain \( D \) are bounded from above or below such that \( Z \) has a finite
(v) the joint density $f$ is continuous and satisfies $f(x, z', 0, V^M) > 0$ for all $x \in D$, then the 2-DEA estimators are statistically consistent in the following sense:

$$\lim_{n \to \infty} \phi_i^{\text{DEA}}(x_i) = \phi(x_i) \cdot \exp(V^M + \zeta) \quad \text{for all } i = 1, \ldots, n. \quad (6)$$

$$\lim_{n \to \infty} \delta^{\text{DEA}} = \delta \quad (7)$$

This result generalizes Proposition 1 by Banker and Natarajan (2008) by relaxing the following two restrictive assumptions: (1) inputs and contextual variables are statistically independent, (2) the effect of contextual variables is one-sided: $Z \geq 0, x \leq 0$. Secondly, we have proved consistency of both the DEA frontier $\phi^{\text{DEA}}$ and the second-stage coefficients $\delta^{\text{DEA}}$. Note that the DEA frontier does not converge to the true frontier $\phi$, it converges to $\phi(x) \cdot \exp(V^M + \zeta)$ (i.e., the frontier augmented by the maximum noise $V^M$ under the ideal conditions represented by $z^x$). Banker and Natarajan (2008) state that $\phi^{\text{DEA}}$ is consistent, but the statement is incorrect if the bias due to the noise term is not taken into account. In practice, the DEA frontier estimator $\phi^{\text{DEA}}$ should be adjusted by factor $\exp(V^M + \zeta)$, which requires that the upper bounds $V^M$ for the noise and $\zeta$ for the contextual effect are known in advance. It seems difficult to estimate $V^M$ or $\zeta$ (or their sum) from data.

Regarding the measurement of efficiency, it is not self-evident how efficiency should be measured in a noisy setting. Define the true output efficiency as $\theta_i = \exp(-u_i)$. Applying Eq. (1), we can state the efficiency measure alternatively as $\theta_i = \frac{|\exp(-u_i) - n|}{|\phi(x_i)|}$. Note that this definition augments the classic Farrell output efficiency measure by taking into account the effect of the contextual variables and noise. The conventional DEA efficiency estimator $\hat{\theta}^{\text{DEA}}$ is inconsistent in the noisy environment because it ignores the effects of contextual variables and random noise on the observed output. Even if the DEA frontier is consistent under truncated noise, the efficiency measure $\hat{\theta}^{\text{DEA}}$ calculated as the radial distance from the observed point to the DEA frontier is inconsistent because the observed data contain noise. Clearly, the statistical consistency of the OLS estimator used in the second stage is not a valid justification for using the conventional DEA routine in the noisy environment.

Compared to the direct OLS estimation of model (2), the main difference in the 2-DEA is that the dependent variable in model (5) is a logarithm of the DEA efficiency estimate rather than the observed output. Note that if the sample covariances of $\ln \phi^{\text{DEA}}(X)$ and all columns of $Z$ are equal to zero, then the OLS estimation of models (2) and (5) yields identical parameter estimates of $\delta$. Thus, if inputs and contextual variables are assumed to be statistically independent, using the DEA method in the first stage appears to provide little benefit compared to the direct OLS estimation of model (2). In fact, Theorem 1 shows that such a strong independence assumption is not required for the consistency of the 2-DEA estimator. If inputs are correlated with contextual variables, the use of the DEA estimator in the first stage can actually help to alleviate the endogeneity problem of the direct OLS estimator. The determination DEA estimator that ignores both the contextual variables and noise is far from perfect, but it may be a better alternative compared to the direct OLS estimator due to the endogeneity problem.

Consistency is a relatively weak property. Intuitively, as the sample size grows, the random sampling procedure will generate an increasing number of “ideal firms” that operate with perfect efficiency under the ideal conditions using the best practices and are lucky enough to be subject to the maximum positive noise. In practice, observing such “ideal firms” is highly unlikely. In the classic deterministic setting ($V^M = 0$), the DEA frontier estimator is known to be downward biased in finite samples (see e.g., Simar and Wilson, 2007). Further, truncation of the dependent variable in Eq. (5) will cause bias for the OLS estimator (note: Theorem 1 shows that this bias converges to zero as the sample size approaches infinity). Interestingly, we can express the bias of the second-stage estimator explicitly as follows:

**Theorem 2.** In finite samples, the bias of the 2-DEA estimator for the coefficients of the contextual variables $(\delta^{\text{2-DEA}})$ depends on the bias of the DEA frontier $\phi^{\text{DEA}}$ as follows:

$$\text{Bias}(\delta^{\text{2-DEA}}) = - (Z')'Z \text{Bias}(\phi^{\text{DEA}}(\mathbf{X})), \quad (8)$$

where

$$\text{Bias}(\phi^{\text{DEA}}(\mathbf{X})) = \left( \begin{array}{c} E[\ln \phi^{\text{DEA}}(\mathbf{x}_1)] - \phi(x_1) \cdot \exp(V^M + \zeta) \\ \vdots \\ E[\ln \phi^{\text{DEA}}(\mathbf{x}_n)] - \ln \phi(x_1) \cdot \exp(V^M + \zeta) \end{array} \right). \quad (9)$$

This theorem shows that the bias of the first-stage DEA estimator (i.e., $\text{Bias}(\phi^{\text{DEA}}(\mathbf{X}))$) will carry over to the second-stage OLS regression. The bias of the DEA estimator is generally not constant for all firms, but depends on the input vector $\mathbf{x}$. Importantly, the bias of the second-stage OLS estimator is due to the correlation of $\text{Bias}(\phi^{\text{DEA}}(\mathbf{X}))$ and $\mathbf{Z}$. The bias of the first-stage DEA estimator is not a problem as such. Note that Eq. (8) can be interpreted as the OLS estimator of the regression equation where one explains $\text{Bias}(\phi^{\text{DEA}}(\mathbf{X}))$ by contextual variables $\mathbf{Z}$. As we assume that inefficiency and noise are independent, the two main causes of correlation of $\text{Bias}(\phi^{\text{DEA}}(\mathbf{X}))$ and $\mathbf{Z}$ are the truncation of the DEA efficiency estimates and the sampling distribution of inputs. We next examine these two main sources of bias separately.

Starting with the truncation of the DEA efficiency estimates, suppose the inputs are randomly drawn from the uniform density over the domain $D$. As random variables $u$ and $v$ are identically and independently distributed, the truncation of the DEA efficiency estimates is the only source of bias. Intuitively, if the true frontier is a smooth concave curve, the bias of the DEA frontier is generally smaller in the vertices of the DEA frontier (i.e., firms classified as efficient in DEA) than in the facets formed as convex combinations of vertices to which the DEA inefficient firms are projected. Since firms that operate under favourable conditions $\mathbf{Z}$ are more likely to be classified as efficient observations in DEA, the bias of the DEA frontier will likely correlate with the contextual variables even if inputs and contextual variables are independently distributed and the inputs are drawn from the uniform density. By Theorem 2, the direction of bias due to truncation is known: the second-stage estimator $\delta^{\text{2-DEA}}$ is biased towards zero. If the true coefficients satisfy $\delta > 0$, the bias of the DEA frontier is positively correlated with the contextual variables (note the negative sign of the bias in (9)), and thus the bias of $\delta^{\text{2-DEA}}$ is negative. Conversely, if $\delta < 0$, the bias of $\delta^{\text{2-DEA}}$ is positive. In fact, it is well known in econometrics that the OLS estimator applied to a truncated dependent variable is biased towards zero (see e.g., Greene, 2008, Chapter 24). However, it seems the known direction of bias due to truncation has been overlooked in the debate concerning the 2-DEA estimator. In the Monte Carlo simulations presented in Section 5, the inputs are drawn from the uniform density. We find that the 2-DEA estimator is systematically biased toward zero in all scenarios considered, even when inputs are independent from the contextual variables.

The situation becomes more complicated if the sampling distribution of inputs is not uniform. The sampling distribution matters because the bias of the DEA frontier tends to be smaller in those regions of the DEA frontier that have a large density of observations.
Possible correlation between inputs and contextual variables complicates matters further. In this general case, the direction of bias is difficult to predict. However, the bias of the DEA frontier could be estimated by applying the bootstrap method (see, e.g., Simar and Wilson, 2007, and references therein). Having estimated the bias of the DEA frontier, we may apply Eq. (8) to estimate the direction and magnitude of the bias of the second-stage estimator \( \hat{\sigma}^{2} \) by simply regressing the estimated bias on the contextual variables \( Z \) using OLS.

In conclusion, Theorem 1 shows that possible correlation of inputs and contextual variables does not influence the statistical consistency of 2-DEA estimator as long as the columns of \( X \) and \( Z \) matrices are not linearly dependent. Theorem 2 presents an explicit formula for the bias in finite samples, which shows that the bias of the DEA frontier in the first-stage carries over to the second-stage OLS estimator through the correlation with the contextual variables. We note that statistical independence of inputs and contextual variables does not necessarily guarantee that \( \text{Bias}(\hat{\sigma}^{2}^{\text{DEA}}|X) \) is uncorrelated with \( Z \) due to the truncation of the DEA efficiency estimates. However, using the DEA efficiency estimator as the dependent variable avoids the endogeneity problem associated with the direct OLS estimation of Eq. (2).

In other words, 2-DEA has an advantage over the direct regression of \( y \) on \( Z \) particularly when inputs are correlated with contextual variables. The Monte Carlo simulations reported in Section 5 support this conclusion. The Monte Carlo evidence presented by Banker and Natarajan further shows that the performance of the 2-DEA estimator deteriorates as the correlation of inputs and contextual variables increases. This observation motivates us to consider another estimation strategy.

4. One-stage semi-parametric estimators

DEA has long been held as fundamentally different from regression analysis and regression-based techniques such as SFA (see e.g., Cooper et al., 2007). Interestingly, the recent study by Kuosmanen and Johnson (2010) shows that the standard output-oriented DEA problem can be equivalently stated as a nonparametric least squares regression subject to monotonicity and concavity constraints on the frontier and sign constraints on the regression parameters just by one, with no impact on the number of known parameters by \( \hat{n} \). The objective function of (10) is globally convex, and the constraints are well-behaved (except for (10a), all constraints are linear).

Consider the 2-DEA method as a combination of two optimization problems: first solve the standard DEA problem involving inputs \( X \) and output \( y \) to obtain the efficiency scores \( h \), and second, solve the least squares problem underlying the OLS estimator where the logarithm of \( h \) is explained by the contextual variables \( Z \). The obvious shortcoming of this procedure is that the impact of the contextual variables \( Z \) is not taken into account in the first stage DEA problem. This problem has been recognized in the SFA literature, where the standard approach nowadays is to integrate the two optimization problems into one problem that jointly estimates the frontier and the impacts of the contextual variables (e.g., Wang and Schmidt, 2002). Parallel to the standard SFA approach, we propose to combine the DEA estimator of the production function and the OLS estimator of \( \sigma^{2} \) into the following nonlinear optimization problem, referred to as 1-DEA problem:

\[
\min_{\theta, \beta, \epsilon} \sum_{i=1}^{n} \epsilon_i^2 \quad \text{s.t.} \quad \ln \gamma_i = \ln \hat{\theta}_i + \mathbf{z}_i' \beta + \epsilon_i \quad \forall i = 1, \ldots, n (10a) \\
\hat{\theta}_i = \hat{\theta} + \mathbf{x}_i' \beta, \quad \forall i = 1, \ldots, n (10b) \\
\mathbf{z}_i' \beta \geq 0 \quad \forall i = 1, \ldots, n (10c) \\
\epsilon_i \leq v^M \quad \forall i = 1, \ldots, n (10d)
\]

The rationale of this 1-DEA formulation is the following. The objective function is to minimize the sum of squares of the composite errors \( \epsilon_i \), which include both inefficiency and truncated noise. Constraint (10a) is a regression equation analogous to (1). In this equation \( \hat{\theta}_i \) is the nonparametric estimator of the unknown frontier point \( \theta(x) \). Constraint (10b) defines the nonparametric estimators as linear tangent hyperplanes, where the coefficients of the hyperplanes are firm-specific. Constraint (10c) enforces the concavity axiom by applying the Afriat (1967, 1972) inequalities (see Kuosmanen, 2008; Kuosmanen and Johnson, 2010; or Kuosmanen and Kortelainen, in press, for a more detailed discussion).

Constraint (10d) imposes free disposability. Finally, truncation of the noise term is explicitly taken into account in constraint (10e) where the composite error \( \epsilon_i \) is enforced to be less than or equal to the truncation point \( v^M \). If the truncation points differ across firms, firm-specific truncation points \( v^M_i \) can be implemented by modifying constraint (10e) as \( \epsilon_i \leq v^M_i \) for all \( i = 1, \ldots, n \).

The objective function of (10) is globally convex, and the constraints are well-behaved (except for (10a), all constraints are linear). Thus problem (10) can be solved by standard nonlinear programming (NLP) algorithms and solvers. NLP solvers are available for example in such mathematical programming packages as GAMS, AIMMS, Matlab, and Lindo, among others. The standard solvers work well in the usual range of sample sizes reported in the literature. However, the computational burden increases at a quadratic rate as the sample size increases. Note that adding a new firm to the sample increases the number of unknown parameters by \( m + 1 \), and the number of constraints increases by \( 2n + 5 \). Introducing an additional input variable increases the number of unknown parameters by \( n \), but there is no impact on the number of constraints. Adding a contextual variable increases the number of parameters just by one, with no impact on the number of constraints.

The 1-DEA formulation presented in (10) has connections to several other models in the recent literature. Firstly, the nonparametric least squares formulation of the standard variable returns to scale DEA problem developed in Kuosmanen and Johnson (2010) is obtained from (10) by simply dropping the contextual variables \( Z \) from (10a) and setting the truncation point equal to zero \( (V^M = 0) \). The logarithmic transformation of the output variable is a monotonic transformation that does not influence the optimal solution. Indeed, the conventional DEA efficiency score of firm \( i \)

---

5 Recognizing that the ordering of the two optimization problems can be arbitrary, Ruggiero (1998) proposes a three-stage approach where the estimated effect of the contextual variables is taken into account in the third-stage DEA efficiency estimation. The main focus of Ruggiero’s study is on the efficiency estimates, and no attempt to correct for the bias in the \( \delta \) coefficients has been made.

6 In our experience, problems with the sample size \( n \leq 100 \) are usually solved in just a few seconds in GAMS. Problems with \( 100 < n < 300 \) usually take from a few minutes to half an hour to solve. The usual sample size in empirical applications is less than 300 observations. Problems with more than 300 observations may take several days, provided that a sufficient amount of physical memory is available. The recent study by Lee et al. (2011) develops more efficient algorithms that are applicable to larger scale problems (e.g., \( n = 600 \)).
is obtained directly as \( \exp(\varepsilon_i) \). Secondly, if the contextual variables \( z \) are excluded from the constraints \((10a)\) and \((10e)\) is relaxed, then the resulting formulation is equivalent to the stochastic semi-nonparametric envelopment of data (StoNED) model described in Kuosmanen and Kortelainen (in press) (see Eq. (30)). Thirdly, if the contextual variables \( z \) are included in the regression equation as stated in \((10a)\) but the truncation constraints \( \varepsilon_i \leq V^M \) \( (i = 1, \ldots, n) \) are excluded, then the resulting formulation is equivalent to stochastic semi-nonparametric envelopment of \( z \)-variables data (StoNEDZ), suggested in the parallel study Johnson and Kuosmanen (2011).

In problem \((10)\), the inputs \( x \) and contextual variables \( z \) are modeled asymmetrically, similar to 2-DEA. In contrast to the firm-specific input coefficients \( x, \beta \), the parameter vector \( \delta \) is common to all firms as in the conventional regression analysis. Although DEA estimators are usually computed by solving a separate LP problem for each firm, a stream of DEA literature integrates the \( n \) LP problems into a single LP problem where the multiplier weights of some inputs/outputs are enforced to be constant across all firms (e.g., Li and Ng, 1995; Kuosmanen et al., 2006). Similarly, our 1-DEA formulation \((10)\) forces the multipliers \( \delta \) to be the same across all firms, while allowing \( x \) and \( \beta \) parameters to vary between firms. On the other hand, if we interpret contextual variables as exogenously fixed inputs (or outputs), in the spirit of Banker and Morey (1986), we could extend the model \((10)\) by introducing firm-specific coefficients \( \delta_i \) that can differ across firms, and include the component \( zx_i \) in the concavity constraints (Afriat inequalities).

In this respect, the 1-DEA formulation \((10)\) can be viewed as a restricted special case of the Banker and Morey model where we impose the constraint \( \delta_i = \delta, (i = 1, \ldots, n) \). When the parameters \( \delta \) are the same for all firms, they can be harmlessly excluded from the concavity constraints.

The statistical properties of the 1-DEA estimator generally depend on the specification of the truncation point \( V^M \) and its possible misspecification. To facilitate a direct comparison of 1-DEA and 2-DEA estimators, we will next examine statistical consistency of the 1-DEA estimator in the special case where the truncation is specified as \( V^M = 0 \) in problem \((10)\), analogous to the conventional DEA estimator.

**Theorem 3.** If the truncation point is specified in problem \((10)\) as \( V^M = 0 \) and the assumptions of Theorem 1 hold, then the 1-DEA estimator is consistent in the same sense as the 2-DEA estimator, specifically:

\[
\lim_{n \to \infty} \hat{\delta}_{1-DEA} = \phi(x_i) \cdot \exp(V^M + \zeta) \quad \text{for all } i = 1, \ldots, n, \quad (11)
\]

\[
\lim_{n \to \infty} \hat{\delta}_{1-DEA} = \delta \quad (12)
\]

The special case \( V^M = 0 \) is interesting because it corresponds to the classic production model with no noise. However, note that the true DGP may include truncated noise; we here consider a situation where the noise term is assumed away in the 1-DEA estimator, which may be a misspecification. In this case, the statistical consistency of 1-DEA and 2-DEA estimators relies essentially on the same set of assumptions regarding the DGP. It is worth noting that the main rationale for modeling the truncation point \( V^M \) explicitly in the 1-DEA problem \((10)\) builds on the assumption that some meaningful information about \( V^M \) is available to the researcher. Otherwise the truncation constraints could be simply excluded. Thus we investigate both 1-DEA with a truncation point and 1-DEA without a truncation point in the next section. However, Theorem 3 suggests that the consistency of 1-DEA need not critically depend on this information. We have shown that the 1-DEA estimator is consistent even if the truncation point \( V^M \) is wrongly specified as zero. Another interpretation of this result is that the explicit modeling of the truncated noise might not provide any benefit when the sample size is very large. However, the truncation does have an impact in relatively small samples. In the next section we will examine how the specification of truncation point and its possible misspecification affect the precision by resorting to Monte Carlo simulations.

## 5. Monte Carlo simulations

This section examines the finite sample performance of the methods discussed in Sections 2–4 in the controlled environment of Monte Carlo simulations. Specifically, we compare the performance of three methods: (1) direct OLS regression of output on contextual variables, (2) conventional 2-DEA,\(^7\) (3) and the proposed 1-DEA method using three alternative specifications of the truncation point \( V^M \). We consider the two extreme cases: \( V^M = 0 \) that is equivalent to the deterministic model with no noise, and \( V^M \to \infty \) that is equivalent to the case of unbounded noise (which has been examined in the parallel study Johnson and Kuosmanen, 2011). In addition to these extreme cases, we also examine the intermediate case where the truncation point is set at one standard deviation from the mean (implying \( V^M = 1.04 \)). We first describe the DGP and the performance statistics used in the simulations, and then report and discuss the results.

### 5.1. Design of experiments

To ensure comparability with earlier Monte Carlo evidence, we replicate the main features of the DGP used by Banker and Natarajan (2008), which was also applied in the parallel study Johnson and Kuosmanen (2011). The production model is the following:

\[
y_i = (x_i^2 - 12z_i^2 + 48z_i - 37) \cdot \exp(2z \delta - u_i + v_i), \quad i = 1, \ldots, n
\]

The true production function is a third-order polynomial of a single input variable \( x \). This function is continuous, monotonic increasing, and concave over the relevant range of inputs used in the simulations. The inputs \( x \) are randomly sampled from a uniform distribution over \([1, 4]\).

Following Wang and Schmidt (2002), we assume that the contextual variables are dependent on the input levels, and are generated according to equation

\[
z = \rho[(x - 1)/3] + w \sqrt{1 - \rho^2}
\]

where parameter \( \rho \) is some pre-specified correlation coefficient and \( w \) is an independent random number drawn from a uniform distribution over the interval \([0, 1]\). The values of \( \rho \in \{-0.8, -0.6, -0.4, -0.2, 0.2, 0.4, 0.6, 0.8\} \) are considered. Note that when \( \rho = 0 \), the input \( x \) and the contextual variable \( z \) are statistically independent. Following Banker and Natarajan (2008), the baseline value of the coefficient \( \delta \) is set as \( \delta = -0.2 \), but the case of a stronger signal with \( \delta = -0.4 \) is also considered.

Random variables \( v \) and \( u \) are drawn independently from \( x, z \), and each other. The inefficiency term \( u \) is drawn from the half-normal density \( |N(0, \sigma_u^2)| \), and the parameter \( \sigma_u \) is set at \( \sigma_u = 0.15 \) in all scenarios considered. In the first four scenarios the noise term \( v \) is drawn from the normal density \( N(0, \sigma_v^2) \). Values of the standard deviation \( \sigma_v \) vary across different scenarios. In the last scenario, the noise term is drawn from the double-truncated normal density with the symmetric truncation at points \( -\tau_v \) and \( \tau_v \).

---

\(^7\) 2-DEA is implemented by estimating the output-oriented variable returns to scale DEA model and regressing the logged DEA efficiency estimates on contextual variables by OLS.
We focus on evaluating performance of alternative methods in the estimation of the coefficient $\delta$, which represents the marginal effect of the contextual variable. We measure performance by the root mean squared deviation (RMSD), defined for the estimator $\hat{\delta}$ as

$$\text{RMSD}(\hat{\delta}) = \frac{100}{\delta} \times \sqrt{\frac{1}{M} \sum_{i=1}^{M} (\hat{\delta}_i - \delta)^2}$$

(15)

where $M$ is the number of simulation trials. The RMSD is always greater than or equal to zero: smaller values for RMSD indicate good performance. Each scenario was replicated 100 times (i.e., $M = 100$). The sample size is $n = 100$ throughout all scenarios examined below.

Additional information about the performance of the methods in terms of bias and mean absolute deviation are provided in the Supplementary Material. In addition to the five most interesting scenarios reported below, we have examined a number of alternative scenarios involving different sample sizes $n$ and parameter values of the DGP. Detailed results of the alternative scenarios are also provided in the Supplementary Material.

5.2. Results

Consider first the base case of Banker and Natarajan where the parameter values are set as $\delta = -0.2, \sigma_x = 0.15, \sigma_z = 0.04, n = 100$. Table 1 reports the RMSD statistics for the three methods and two alternative truncation points for 1-DEA while varying the levels of the correlation coefficient $\rho$.

The direct OLS estimator performs extremely poorly even when $\rho = 0$ indicating the inputs and the contextual variable are statistically independent. Poor performance of the direct OLS estimator is due to the endogeneity problem discussed in Section 2.1. The endogeneity problem is further aggravated by the fact that the variance of the composite error term is very large if the input variable $x$ is omitted. Even though the 2-DEA estimator is not the most competitive method in Table 1, our results demonstrate that the use of DEA in the first stage can help to circumvent the endogeneity problem of the direct OLS estimation.

The results of Table 1 show a major advantage for the proposed 1-DEA estimators. The specification of the truncation point $V^M$ does not have a major impact on performance, and even the wrongly specified 1-DEA estimators are more precise than the conventional 2-DEA estimator. In this baseline scenario the noise term is drawn from the untruncated normal density, but the 1-DEA specification where we set $V^M = 1.04$ has a notably lower RMSD when $\rho = -0.4$, $\delta = -0.2$, and $V^M = \infty$ that does not assume truncation. However, the differences in the RMSD of the alternative 1-DEA specifications are relatively small. The difference relative to 2-DEA is more pronounced, especially when the contextual variable is correlated with the input. Large absolute value of $\rho$ has a negative effect on the performance of the 1-DEA estimators, but the effect is much smaller than in the case of 2-DEA.

We next examine alternative scenarios involving different parameter values to examine how the signal, the noise, and the truncation of the noise term influence the performance of the three methods.

Let us first examine the effect of a stronger signal from the contextual variable. We change the coefficient of contextual variable to $\delta = -0.4$, but keep the other parameters of the baseline scenario unchanged. The contextual variable $z$ has a larger negative effect on output thus it should be easier to identify. Table 2 reports the results for this scenario, organized analogous to Table 1. The main differences can be observed in the precision of the estimated $\delta$ coefficients. The direct OLS estimator improves its performance, but it remains as the weakest method. Performance of 2-DEA improves when the correlation of $x$ and $z$ is low, but there is no improvement when the correlation of $x$ and $z$ is high. In contrast, performance of the 1-DEA estimators improves through all parameter values $\rho$. The relative rankings of the alternative specifications of 1-DEA remain the same as in Table 1.

The noise term obviously influences the precision of the estimators. We examine the impact of the noise term in the next two scenarios. The third scenario we consider is the classical DEA setting where the noise term is excluded. In practice, we set the standard deviation of the noise term equal to zero (i.e., $\sigma_z = 0$), keeping all other parameter values at the same level as in the baseline scenario of Table 1. The results of the no noise scenario are reported in Table 3.

As excluding the noise term decreases the variance of the composite disturbance term, all methods improve their performance compared to the baseline scenario. Note that the correctly specified 1-DEA estimator with $V^M = 0$ achieves the lowest RMSD at all levels of the correlation parameter $\rho$. This scenario shows that imposing the truncation constraint 1-DEA can be beneficial. However, even if the truncation point is wrongly specified, or excluded completely, the 1-DEA estimator performs better than the conventional 2-DEA estimator, especially when the absolute value of the correlation parameter $\rho$ is high.

The fourth scenario is referred to as the heavy noise scenario because we increase the standard deviation of the noise term, $\sigma_z = 0.09$. All other parameter values are maintained at their baseline levels. As the true noise term is unbounded in this scenario, the case $V^M = \infty$ is the correct specification of the truncation point. The results of the heavy noise scenario are reported in Table 4.

As the variance of the composite disturbance term increases, the performance of all methods is expected to deteriorate in this scenario compared to the baseline case of Table 1. However, the 2-DEA estimator performs generally better in this scenario. This is due to the fact that the noise term causes upward bias in the otherwise downward biased DEA estimator, and thus the two sources of bias cancel out. The wrongly specified 1-DEA estimator with $V^M = 0$ does not benefit from this effect, and its performance deteriorates in the heavy noise scenario. The correctly specified untruncated 1-DEA estimator achieves the lowest RMSD when the correlation parameter $\rho$ differs from zero, but the wrongly specified 1-DEA estimator with $V^M = 1.04$ proves also competitive, achieving the lowest RMSD when the correlation parameter $\rho$ is equal to zero.

In the last scenario we examine how the truncation of the true noise term influences performance. We note that Banker and Natarajan (2008) state they draw their noise term from the double-truncated normal density, but as they set the truncation points at

### Table 1

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>$\rho = -0.8$</th>
<th>$\rho = -0.6$</th>
<th>$\rho = -0.4$</th>
<th>$\rho = -0.2$</th>
<th>$\rho = 0.0$</th>
<th>$\rho = 0.2$</th>
<th>$\rho = 0.4$</th>
<th>$\rho = 0.6$</th>
<th>$\rho = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>774</td>
<td>572</td>
<td>404</td>
<td>210</td>
<td>117</td>
<td>249</td>
<td>409</td>
<td>585</td>
<td>753</td>
</tr>
<tr>
<td>2-DEA</td>
<td>74.0</td>
<td>50.0</td>
<td>32.2</td>
<td>22.0</td>
<td>18.9</td>
<td>20.6</td>
<td>24.1</td>
<td>38.4</td>
<td>57.6</td>
</tr>
<tr>
<td>1-DEA, $V^M = 0$</td>
<td>32.3</td>
<td>25.2</td>
<td>23.1</td>
<td>20.9</td>
<td>20.3</td>
<td>20.6</td>
<td>24.3</td>
<td>24.5</td>
<td>32.0</td>
</tr>
<tr>
<td>1-DEA, $V^M = 1.04$</td>
<td>29.5</td>
<td>23.2</td>
<td>19.4</td>
<td>13.4</td>
<td>15.2</td>
<td>19.5</td>
<td>22.5</td>
<td>23.8</td>
<td>30.5</td>
</tr>
<tr>
<td>1-DEA, $V^M = \infty$</td>
<td>29.1</td>
<td>22.1</td>
<td>19.5</td>
<td>18.3</td>
<td>17.9</td>
<td>18.3</td>
<td>17.7</td>
<td>28.4</td>
<td>30.1</td>
</tr>
</tbody>
</table>

Note: baseline scenario: $\delta = -0.2$, untruncated noise with $\sigma_z = 0.04$. 

six standard deviations from the mean (zero), the probability that truncation occurs is astronomically small (approximately $1.97 \times 10^{-9}$). In the last scenario we set the truncation point only one standard deviation from the mean: the noise term is restricted to the interval $[-0.04,0.04]$. In this case, the probability that truncation occurs is 31.7%. The results of the truncated noise scenario are presented in Table 5.

In this scenario the 1-DEA estimator with $V^M = 1.04$ is correctly specified. It achieves the lowest RMSE when the correlation parameter $\rho = 0.2$, but at all other parameter values the wrongly specified 1-DEA estimator with $V^M = 0$ performs better. Note that the standard deviation of the truncated noise term $v$ is equal to 0.0116, not 0.04 (i.e., the parameter value of the untruncated normal density). The low standard deviation of the truncated noise term can at least partly explain why the 1-DEA specification with $V^M = 0$ performs well in this scenario. However, the specification $V^M \rightarrow \infty$ is also relatively competitive.

In conclusion, the proposed 1-DEA method that simultaneously estimates the nonparametric frontier and the coefficients of the contextual variables performs systematically better than the conventional 2-DEA at almost all parameter values considered. To put this result in a proper context, we have also presented the RMSE statistics for the direct OLS estimator, which is statistically consistent if $x$ and $z$ are statistically independent. Based on the available evidence (see the Supplementary Material for alternative scenarios and more detailed performance statistics), we may conclude that 2-DEA remains a competitive alternative when the inputs and contextual variables are uncorrelated. However, the performance of 2-DEA deteriorates rapidly as the correlation of $x$ and $z$ increases (in the absolute sense). For all methods, the correlation of $x$ and $z$ has a negative effect on the precision, but the proposed 1-DEA method proves much more robust to this correlation. Regarding the specification of the 1-DEA estimator, our results demonstrate that imposing the truncation constraint can improve precision in some cases, particularly in the fully deterministic case and when the correlation of $x$ and $z$ is close to zero. However, the exact specification of the truncation point seems not critically important. In many cases a wrongly specified 1-DEA estimator achieved the lowest RMSE.

### 6. Conclusions

Estimating the effects of contextual variables is of critical importance for practitioners of efficiency analysis. This paper has further investigated the performance of several methods to

---

Table 2: Performance (by RMSD) of alternative methods in estimating the effect of contextual variable ($\delta$) under different correlation levels ($\rho$). The high signal scenario.

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>$\rho = -0.8$</th>
<th>$\rho = -0.6$</th>
<th>$\rho = -0.4$</th>
<th>$\rho = -0.2$</th>
<th>$\rho = 0.0$</th>
<th>$\rho = 0.2$</th>
<th>$\rho = 0.4$</th>
<th>$\rho = 0.6$</th>
<th>$\rho = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>379</td>
<td>286</td>
<td>193</td>
<td>116</td>
<td>64.1</td>
<td>112</td>
<td>195</td>
<td>300</td>
<td>383</td>
</tr>
<tr>
<td>2-DEA</td>
<td>78.0</td>
<td>50.9</td>
<td>30.5</td>
<td>17.2</td>
<td>11.3</td>
<td>11.9</td>
<td>13.4</td>
<td>31.4</td>
<td>51.1</td>
</tr>
<tr>
<td>1-DEA, $V^M = 0$</td>
<td>15.0</td>
<td>11.3</td>
<td>9.90</td>
<td>9.26</td>
<td>11.4</td>
<td>10.2</td>
<td>10.0</td>
<td>12.2</td>
<td>15.9</td>
</tr>
<tr>
<td>1-DEA, $V^M = 1.04$</td>
<td>14.8</td>
<td>11.5</td>
<td>8.8</td>
<td>6.7</td>
<td>7.6</td>
<td>9.9</td>
<td>11.2</td>
<td>12.0</td>
<td>15.1</td>
</tr>
<tr>
<td>1-DEA, $V^M \rightarrow \infty$</td>
<td>12.8</td>
<td>9.70</td>
<td>8.52</td>
<td>7.95</td>
<td>9.71</td>
<td>9.38</td>
<td>8.41</td>
<td>11.5</td>
<td>15.0</td>
</tr>
</tbody>
</table>

Note: high signal scenario: $\delta = -0.4$, untruncated noise with $\sigma_x = 0.04$.

Table 3: Performance (by RMSD) of alternative methods in estimating the effect of contextual variable ($\delta$) under different correlation levels ($\rho$). No noise scenario.

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>$\rho = -0.8$</th>
<th>$\rho = -0.6$</th>
<th>$\rho = -0.4$</th>
<th>$\rho = -0.2$</th>
<th>$\rho = 0.0$</th>
<th>$\rho = 0.2$</th>
<th>$\rho = 0.4$</th>
<th>$\rho = 0.6$</th>
<th>$\rho = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>762</td>
<td>572</td>
<td>385</td>
<td>235</td>
<td>143</td>
<td>236</td>
<td>365</td>
<td>585</td>
<td>767</td>
</tr>
<tr>
<td>2-DEA</td>
<td>68.5</td>
<td>44.6</td>
<td>33.9</td>
<td>18.9</td>
<td>17.2</td>
<td>19.2</td>
<td>26.1</td>
<td>39.9</td>
<td>61.4</td>
</tr>
<tr>
<td>1-DEA, $V^M = 0$</td>
<td>15.1</td>
<td>11.3</td>
<td>10.5</td>
<td>9.35</td>
<td>10.4</td>
<td>9.27</td>
<td>9.86</td>
<td>11.3</td>
<td>15.0</td>
</tr>
<tr>
<td>1-DEA, $V^M = 1.04$</td>
<td>25.4</td>
<td>22.3</td>
<td>21.7</td>
<td>13.8</td>
<td>13.9</td>
<td>17.1</td>
<td>20.6</td>
<td>22.1</td>
<td>28.3</td>
</tr>
<tr>
<td>1-DEA, $V^M \rightarrow \infty$</td>
<td>26.2</td>
<td>19.7</td>
<td>18.5</td>
<td>16.4</td>
<td>16.5</td>
<td>17.5</td>
<td>20.0</td>
<td>26.8</td>
<td></td>
</tr>
</tbody>
</table>

Note: no noise scenario: $\delta = -0.2$, untruncated noise with $\sigma_x = 0$.  

Table 4: Performance (by RMSD) of alternative methods in estimating the effect of contextual variable ($\delta$) under different correlation levels ($\rho$). Heavy noise scenario.

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>$\rho = -0.8$</th>
<th>$\rho = -0.6$</th>
<th>$\rho = -0.4$</th>
<th>$\rho = -0.2$</th>
<th>$\rho = 0.0$</th>
<th>$\rho = 0.2$</th>
<th>$\rho = 0.4$</th>
<th>$\rho = 0.6$</th>
<th>$\rho = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>780</td>
<td>562</td>
<td>386</td>
<td>229</td>
<td>141</td>
<td>233</td>
<td>387</td>
<td>554</td>
<td>760</td>
</tr>
<tr>
<td>2-DEA</td>
<td>86.0</td>
<td>58.1</td>
<td>38.0</td>
<td>26.1</td>
<td>25.1</td>
<td>27.3</td>
<td>28.5</td>
<td>36.8</td>
<td>50.6</td>
</tr>
<tr>
<td>1-DEA, $V^M = 0$</td>
<td>61.1</td>
<td>47.8</td>
<td>42.5</td>
<td>40.3</td>
<td>46.0</td>
<td>42.6</td>
<td>48.1</td>
<td>63.4</td>
<td></td>
</tr>
<tr>
<td>1-DEA, $V^M = 1.04$</td>
<td>39.8</td>
<td>29.2</td>
<td>23.6</td>
<td>16.9</td>
<td>20.2</td>
<td>25.2</td>
<td>28.0</td>
<td>32.4</td>
<td>38.3</td>
</tr>
<tr>
<td>1-DEA, $V^M \rightarrow \infty$</td>
<td>34.3</td>
<td>26.3</td>
<td>23.3</td>
<td>21.9</td>
<td>24.7</td>
<td>25.2</td>
<td>26.3</td>
<td>27.1</td>
<td>35.8</td>
</tr>
</tbody>
</table>

Note: heavy noise scenario: $\delta = -0.2$, untruncated noise with $\sigma_x = 0.09$.

Table 5: Performance (by RMSD) of alternative methods in estimating the effect of contextual variable ($\delta$) under different correlation levels ($\rho$). Truncated noise scenario.

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>$\rho = -0.8$</th>
<th>$\rho = -0.6$</th>
<th>$\rho = -0.4$</th>
<th>$\rho = -0.2$</th>
<th>$\rho = 0.0$</th>
<th>$\rho = 0.2$</th>
<th>$\rho = 0.4$</th>
<th>$\rho = 0.6$</th>
<th>$\rho = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>764</td>
<td>564</td>
<td>393</td>
<td>224</td>
<td>132</td>
<td>241</td>
<td>378</td>
<td>581</td>
<td>758</td>
</tr>
<tr>
<td>2-DEA</td>
<td>84.8</td>
<td>56.7</td>
<td>35.2</td>
<td>22.0</td>
<td>17.4</td>
<td>17.6</td>
<td>19.4</td>
<td>25.4</td>
<td>41.4</td>
</tr>
<tr>
<td>1-DEA, $V^M = 0$</td>
<td>24.4</td>
<td>17.5</td>
<td>15.7</td>
<td>14.6</td>
<td>16.0</td>
<td>16.5</td>
<td>17.5</td>
<td>18.3</td>
<td>24.2</td>
</tr>
<tr>
<td>1-DEA, $V^M = 1.04$</td>
<td>27.9</td>
<td>21.4</td>
<td>19.5</td>
<td>17.2</td>
<td>16.3</td>
<td>16.4</td>
<td>18.2</td>
<td>20.7</td>
<td>26.9</td>
</tr>
<tr>
<td>1-DEA, $V^M \rightarrow \infty$</td>
<td>27.9</td>
<td>24.0</td>
<td>18.4</td>
<td>17.3</td>
<td>16.3</td>
<td>16.9</td>
<td>18.2</td>
<td>20.7</td>
<td>26.8</td>
</tr>
</tbody>
</table>

Note: truncated noise scenario: $\delta = -0.2$, double-truncated noise with $\sigma_x = 0.04$, $\tau_x = 0.04$.
estimate the effect of contextual variables. A new one-stage estimation strategy is proposed that facilitates joint estimation of an axiomatic production function with a linear regression model for the contextual variables.

To complement the asymptotic results, Monte Carlo simulations were conducted to investigate the finite sample performance of the developed methods. The evidence from these simulations shows that the proposed approaches outperform conventional 2-DEA when the contextual variables are correlated with inputs. We find that the 1-DEA method yields the most precise estimates when no noise is present. In the noisy environments, the bias in 2-DEA helps to offset the effects of the small sample bias of DEA and thus outperforms 1-DEA.

We suggest that our findings can provide insight into the relationship between estimating efficiency and modeling the context of production. Modeling these two components will assist managers who develop business strategies or determine operational practices, and policy-makers charged with writing regulations. Our approach is applicable to a variety of industries and levels of analysis. Extending this paper’s results to a multi-output setting is a logical next step for future research.

Appendix A. Proofs of theorems

Theorem 1. Greene (2008, Definition D.2) defines statistical consistency in terms of the probability limit operator plim, as stated in Eqs. (6) and (7). Equivalently, we can state consistency using the standard definition of convergence in probability (Greene, 2008, Definitions D.1 and D.6):

\[ \lim_{n \to \infty} \Pr(\hat{\phi}_{\text{DEA}}^n(x) - \phi(x) \cdot \exp(VM^t + \zeta_i) > \varepsilon) = 0 \]

for all \( i = 1, \ldots, n \). (16)

\[ \lim_{n \to \infty} \Pr(\hat{\phi}_{\text{DEA}}^n(x) - \phi(x) > \varepsilon) = 0 \]

(17)

where \( \varepsilon > 0 \) denotes an arbitrarily small positive constant.

Consider first the nonparametric part \( \hat{\phi}_{\text{DEA}} \), defined in (3). The logic of the first part of the proof is analogous to that of Proposition 5 by Banker (1993), which established consistency of the classic DEA estimator. Note that the true production function \( \phi(x) \) is assumed to be concave, and hence continuous. Thus, function \( \phi(x) \cdot \exp(VM^t + \zeta_i) \) is also continuous.

To gain intuition, consider an arbitrary randomly drawn observation \((y_i, x_i)\). Suppose our random sample contains a reference firm \( k \) that satisfies the following equations:

\[ x_k = x_i, \quad z_k = z_i, \quad u_{ik} = 0, \quad v_k = VM^t. \]

Hence, \( y_k = \phi(x_i) \cdot \exp(VM^t + \zeta_i) \). The values of \( x_k, z_k, u_{ik} \), and \( v_k \) have been chosen such that it is impossible to achieve a higher observed output level than \( y_k \) by using the input vector \( x_i \). Thus, if the observation \( k \) characterized by (18) is contained in the observed sample, then the DEA estimator of the production function is simply \( \hat{\phi}_{\text{DEA}}(x_i) = y_k \). Otherwise, \( \hat{\phi}_{\text{DEA}}(x_i) \leq y_k \). Note that if \( \phi \) is a twice continuously differentiable function, then \( \hat{\phi}_{\text{DEA}}(x_i) < y_k \) whenever the observation \( k \) that satisfies (18) is not drawn to the sample.

Since the joint distribution of the random variables \((x, z, u, v)\) is continuous by assumption, the probability of observing firm \( k \) that satisfies equalities (18) is equal to zero. However, assumption (v) implies there is a positive probability of observing a firm in a neighborhood of firm \( k \). Specifically, for any arbitrary input level \( x_i \), there is a positive probability of observing a firm \( h \) that satisfies

\[ x_h \in [x_i - \Delta 1, x_i], \quad z_h \in [z_i - \Delta 1, z_i], \quad u_h \in [0, \Delta], \quad v_h \in [VM^t - \Delta, VM^t], \]

where \( \Delta > 0 \) is an arbitrary positive constant. For any given \( \varepsilon > 0 \), it is possible to find a constant \( \Delta > 0 \) and probability \( p_i > 0 \) such that

\[ \Pr(\phi(x) > \varepsilon) = \exp(VM^t + \zeta - \phi(x_i) \cdot \exp(VM^t + \zeta)) > \varepsilon \]

If firm \( h \) that satisfies (19) is observed, the maximum error of the DEA estimator is equal to \( \varepsilon \). However, the error can be smaller. Further, it is possible that the error is smaller than \( \varepsilon \) even if a firm \( h \) satisfying conditions (19) is not observed. Thus, the following inequality applies:

\[ \Pr(\phi_{\text{DEA}}^n(x) - \phi(x) \cdot \exp(VM^t + \zeta) > \varepsilon) \leq p_i. \]

In general, consistency of \( \hat{\phi}_{\text{DEA}} \) requires that, for any arbitrary \( \varepsilon > 0 \), the probability of drawing firm \( h \) to the sample approaches unity as the sample size approaches infinity. The probability that firm \( h \) is not observed in a sequence of \( n \) independent random draws is equal to \((1 - p_i)^n\). Asymptotically, this probability converges to zero: \( \lim_{n \to \infty} (1 - p_i)^n = 0 \). Hence,

\[ \lim_{n \to \infty} \Pr(\phi_{\text{DEA}}^n(x) - \phi(x) \cdot \exp(VM^t + \zeta) > \varepsilon) = 0. \]

This is equivalent to \( \lim_{n \to \infty} \hat{\phi}_{\text{DEA}}^n(x) = \phi(x) \cdot \exp(VM^t + \zeta) \). As the argument was made for an arbitrary \( x_i \), the same argument applies to all observations \( i = 1, \ldots, n \).

Next, consider the parametric part. Recall that \( \ln \theta_i = \ln y_i - \ln \hat{\phi}_{\text{DEA}}(x_i) \). Form the column vector \( \ln \theta = (\ln \theta_1, \ldots, \ln \theta_n) \) that represents the dependent variable of the second-stage OLS problem. Note that the OLS estimator has the well-known closed-form solution:

\[ \delta_{Z}^{2-\text{DEA}} = (ZZ)^{-1}Z[\ln \theta] \]

Denote \( \eta = \left(\eta_1, \ldots, \eta_n\right) \), where \( \eta_i = \ln \hat{\phi}_{\text{DEA}}(x_i) - (\ln \phi(x_i) + VM^t + \zeta) \) is the finite sample error of the DEA frontier evaluated at the input level of firm \( i \). Applying this notation, we can develop the standard OLS formula as

\[ \delta_{Z}^{2-\text{DEA}} = (ZZ)^{-1}Z[\ln \theta] \]

First, note that since \((VM + \zeta)\) is a constant scalar,

\[ (ZZ)^{-1}Z[VM + \zeta] = (VM + \zeta)(ZZ)^{-1}Z = 0. \]

Next, since the noise term is assumed to be statistically independent of the contextual variables, we have \( E[(ZZ)^{-1}Z\nu] = 0 \). Applying the generalized Chebychev’s inequality (Greene, 2008, Theorem D.10),

\[ \Pr\left(\|ZZ^{-1}Z\nu\| < \varepsilon\right) \leq \frac{\varepsilon^2}{\|ZZ^{-1}Z\nu\|^2} = 0. \]

Thus, we have \( \lim_{n \to \infty} (ZZ)^{-1}Z\nu = 0. \)
Applying the previous argument to the inefficiency term, we also have $E[(ZZ)^{-1}Zu] = 0$, which implies $\lim_{n \to \infty} E[(ZZ)^{-1}Z] = 0$.

Finally, consistency of the DEA estimator $\hat{\delta}^{\text{DEA}}(x_k)$ established in the first part of this proof implies that $\lim_{n \to \infty} E[(ZZ)^{-1}Z] = 0$.

We find that the effects of the noise terms $v$, inefficiency terms $u$, the constant bound $(V^M + \zeta)$, and the finite sample error $\eta$ all vanish asymptotically. Thus, $\lim_{n \to \infty} \hat{\delta}^{\text{DEA}} = \delta$. □

**Theorem 2.** From the proof of Theorem 1, we have (20):
\[
\hat{\delta}^{\text{DEA}} = \delta + (ZZ)^{-1}Z(E(v) - E(u) - (V^M + \zeta)1 - E(\eta)).
\]

We have used the fact that $E(u) = E(1) = \mu 1$ and $E(v) = 0$. Next, note that $E(\eta) = \text{Bias}(\hat{\delta}^{\text{DEA}}(X))$. Recognizing $(\mu + V^M + \zeta)$ as a constant and $(ZZ)^{-1}Z$ as the OLS formula for the regression model where a vector of ones is regressed on the $z$-variables, we have $-(ZZ)^{-1}Z(\mu + V^M + \zeta)1 = -(\mu + V^M + \zeta)(ZZ)^{-1}Z1 = 0$. Thus, the bias of $\hat{\delta}^{\text{DEA}}$ is
\[
\text{Bias}(\hat{\delta}^{\text{DEA}}) = E[(ZZ)^{-1}Z(E(v) - E(u) - (V^M + \zeta)1 - E(\eta))].
\]

**Theorem 3.** Consider the restricted special case of the 1-DEA problem (10) obtained by imposing the following parameter restrictions: $\delta = 0$ and $V^M = 0$. Note that these restrictions are imposed to the programming problem (10), but the true coefficients $\delta$ and the truncation point $V^M$ need not satisfy these restrictions if the model represented by (10) is incorrectly specified. The restricted version of the 1-DEA problem (10) can be stated as
\[
\min_{\mathbf{z}, \mathbf{\phi}, \mathbf{\beta}} \sum_{i=1}^{n} c_i^2 \quad \text{s.t.} \quad \begin{align*}
\ln y_i &= \ln \phi_i + c_i & \text{for all } i = 1, \ldots, n \\
\phi_i &= u_i + x_i \beta_i & \text{for all } i = 1, \ldots, n \\
\phi_i &\leq x_i + x_i \beta_i & \text{for all } i = 1, \ldots, n \\
\beta_i &\geq 0 & \text{for all } i = 1, \ldots, n \\
c_i &\leq 0 & \text{for all } i = 1, \ldots, n 
\end{align*}
\]

Problem (21) is a variant of the sign-constrained nonparametric least-squares problem, which Kuosmanen and Johnson (2010) have shown to be equivalent to the standard output-oriented DEA estimator under VRS. The result is based on the fact that quadratic objective function and the logarithm transformation are both monotonic transformations that do not influence the shape of the DEA frontier. Thus, the estimator $\hat{\delta}^{\text{DEA}}(x_k)$ obtained from the reduced 1-DEA problem (21) is equivalent to the classic DEA estimator $\hat{\delta}^{\text{DEA}}$.

Consistency of the DEA estimator $\hat{\delta}^{\text{DEA}}(x_k)$ was shown in Theorem 1, and the same arguments can be directly applied to the reduced 1-DEA estimator $\hat{\delta}^{\text{DEA}}$. Note that the DEA estimator is consistent even when the contextual variables are omitted completely: all three sources of deviations from $\phi(x_k) \cdot \exp(V^M + \zeta)$ have a negative sign. The inefficiency term $u_i$ has an explicit sign-constraint. The noise term $\eta$ is assumed to be right-truncated at $V^M$, but by Theorem 1, the DEA estimator $\hat{\delta}^{\text{DEA}}$ is consistent even if the restriction $V^M = 0$ does not hold. Finally, the effect of the contextual variables is assumed to be less than or equal to $\zeta$. Hence, relaxing the constraint $\delta = 0$ does not have any influence on the estimated one-stage DEA frontier: recall that in DEA the observations below the frontier have no effect on the shape of the DEA frontier.

To prove consistency of the coefficients of the contextual variables, we denote $\ln y_i^{\text{DEA}} = \ln y_i - \ln \hat{\delta}^{\text{DEA}}$ and form a column vector $\ln \theta^{\text{DEA}} = (\ln \theta^{\text{DEA}} - \ln \theta^{\text{DEA}})$. Without a loss of generality, we can focus on the subproblem of determining coefficients $\hat{\delta}^{\text{DEA}}$ conditional on $\ln \theta^{\text{DEA}}$,
\[
\min_{\mathbf{d}} (\ln \theta^{\text{DEA}} - Z \delta)(\ln \theta^{\text{DEA}} - Z \delta)(\ln \theta^{\text{DEA}} - Z \delta) = 0.
\]

The Lagrangian of this problem is
\[
L = (\ln \theta^{\text{DEA}} - Z \delta)(\ln \theta^{\text{DEA}} - Z \delta) - \lambda(\ln \theta^{\text{DEA}} - Z \delta),
\]
where $\lambda$ is a vector of Lagrange multipliers. Differentiating $L$ with respect to $\delta$ yields the first-order conditions
\[
Z(\ln \theta^{\text{DEA}} - Z \delta) = \lambda Z = 0.
\]

Solving $\delta$ from this system of equations yields a unique closed-form solution, analogous to the standard OLS formula:
\[
\hat{\delta}^{\text{DEA}} = (ZZ)^{-1}Z(\ln \theta^{\text{DEA}} - Z \delta).
\]

Next, note that the consistency of $\hat{\delta}^{\text{DEA}}$ implies that the DEA efficiency measure $\ln \theta^{\text{DEA}}$ also converges in probability. Note that $\ln \theta^{\text{DEA}}$ does not converge to a constant; it converges to a random variable, which is the sum of random noise, inefficiency, and the effect of contextual variables. The convergence in probability can be stated in terms of the probability limit operator as follows:
\[
\lim_{n \to \infty} \ln \theta^{\text{DEA}} = (Z \delta - \zeta) + (V_i - V^M) - u_i.
\]

This will further imply that,
\[
\lim_{n \to \infty} \ln \theta^{\text{DEA}} - Z \delta = -\zeta + (V_i - V^M) - u_i \leq 0.
\]

This inequality reveals that the envelopment constraint $\ln \theta^{\text{DEA}} - Z \delta \leq 0$ becomes non-binding as the sample size approaches infinity. Therefore, as the constraints are asymptotically redundant, the optimal solution to the constrained optimization problem (22) converges in probability to the optimal solution of the unconstrained least squares problem:
\[
\lim_{n \to \infty} \ln \theta^{\text{DEA}} = (Z \delta - \zeta).
\]

Finally, we can apply the arguments presented in the second part of Theorem 1 to verify that
\[
\lim_{n \to \infty} \hat{\delta}^{\text{DEA}} = \delta. \quad \square
\]

**Appendix B. Supplementary Material**

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2012.01.023.

**References**


