CEFAGE-UE Working Paper
2010/01

Fractional Regression Models for Second Stage DEA Efficiency Analyses

Esmeralda A. Ramalho, Joaquim J.S.Ramalho, Pedro D. Henriques

Departamento de Economia, Universidade de Évora and CEFAGE-UE

Fractional regression models for second stage DEA efficiency analyses*

Esmeralda A. Ramalho, Joaquim J.S.Ramalho and Pedro D. Henriques

Departamento de Economia and CEFAGE-UE, Universidade de Évora

This Draft: January 2010

Abstract

Data envelopment analysis (DEA) is commonly used to measure the relative efficiency of decision-making units. Often, in a second stage, a regression model is estimated to relate DEA efficiency scores to exogenous factors. In this paper, we argue that the traditional linear or tobit approaches to second-stage DEA analysis do not constitute a reasonable data-generating process for DEA scores. Under the assumption that DEA scores can be treated as descriptive measures of the relative performance of units in the sample, we show that using fractional regression models are the most natural way of modeling bounded, proportional response variables such as DEA scores. We also propose generalizations of these models and, given that DEA scores take frequently the value of unity, examine the use of two-part models in this framework. Several tests suitable for assessing the specification of each alternative model are also discussed.

Keywords: Second-stage DEA, fractional data, specification tests, one outcomes, two-part models.

JEL Classification: C12, C13, C25, C51.

*Financial support from Fundação para a Ciência e a Tecnologia is gratefully acknowledged (grant PTDC/ECO/64693/2006). Address for correspondence: Esmeralda A. Ramalho, Department of Economics, Universidade de Évora, Largo dos Colegiais, 7000-803 ÉVORA, Portugal (e-mail: ela@uevora.pt).

1 Introduction

Data envelopment analysis (DEA) is a technique widely used to evaluate the relative efficiency of individual decision-making units (DMUs). DEA efficiency scores (y) are typically defined on the interval $]0,1]$, with, in general, few values, if any, close to 0 but some values of unity. In order to examine the effect on the efficiency of DMUs of factors that are beyond their control (the so-called environmental, contextual or non-discretionary variables), often, in a second stage, a regression model is estimated for DEA scores.

The choice of regression model for the second stage of DEA analysis is not a trivial econometric problem. The standard linear model is not, in general, appropriate for such analysis, since the predicted values of y may lie outside the unit interval and the implied constant marginal effects of the covariates on y are not compatible with both the bounded nature of DEA scores and the existence of a mass point at unity in their distribution. Moreover, the standard approach of using a two-limit tobit model, with limits at zero and unity, to model DEA scores (see the references in Simar and Wilson 2007) is also questionable. Indeed, the accumulation of observations at unity is a natural consequence of the way DEA scores are defined rather than the result of censoring. Furthermore, the domain of the two-limit tobit model differs from that of DEA scores because typically efficiency scores of zero are not observed. This difference is particularly relevant because application of the two-limit tobit model in this context in fact amounts to estimation of a one-limit tobit for $y \in]-\infty, 1]$.

In the second-stage DEA literature, despite the acknowledgement by authors such as Ruggiero (1998) that misspecifying the second-stage regression model may generate misleading results, only recently has this issue been addressed. Given the bounded nature of DEA scores, both Hoff (2007) and McDonald (2009) considered the use of Papke and Wooldridge's (1996) logit fractional regression model. However, both researchers conclude by recommending the use of the simpler linear regression model, although McDonald (2009) acknowledges that there are advantages in using Papke and Wooldridge's (1996) model for more refined analyses.

In this paper, we argue that any sensible description of the data-generating process (DGP) for DEA scores defined on $]0,1]$ requires the use of regression models that are appropriate for dealing with fractional data in the second-stage DEA analysis. In contrast to Hoff (2007) and McDonald (2009), who considered only the logit fractional regression model, we analyze several alternative functional forms that may be more useful for dealing with the typical asymmetric nature of DEA scores. In addition, we provide examples to show clearly that both linear and tobit models may produce outcomes that differ greatly from those of fractional regression models (FRMs). Another advantage of using FRMs is that they may be estimated by quasi-maximum likelihood (QML). This is because, unlike tobit models, FRMs do not require assumptions to be made about the conditional distribution of DEA scores or heteroskedasticity patterns.

The relatively high proportion of efficient DMUs usually found in empirical studies raises another issue in the description of the DGP for DEA scores: should the values of unity be treated differently? In fact, there are two options: (i) use a single-equation model to explain the DEA scores of all DMUs, including those of the efficient ones; or (ii) use a two-equation model that explains separately, first, why some DMUs are efficient while others are not ($y = 1$ versus $y < 1$) and, second, the relative efficiency of inefficient DMUs. To our knowledge, most researchers use one-part models, an exception being Hoff (2007). In this paper, we discuss two-part models that are more flexible than those proposed by Hoff (2007). This greater flexibility derives from our use of QML (rather than maximum likelihood) for estimation of the second component of these models, because QML enables one to avoid the distributional assumptions made by Hoff (2007).

The validity of the assumptions that underlie the models used in the DEA second-stage regressions has been systematically overlooked in the DEA literature: most empirical practitioners do not test the conditional mean and/or the distributional assumptions implied by their models. Because FRMs estimated by QML merely require the correct specification of the conditional mean of the DEA scores, in this paper we focus on this issue and survey some general specification tests that can be applied in this framework.

The paper is organized as follows. In Section 2, we establish the framework of the paper. In Section 3, we examine alternative regression models for DEA scores, including the traditional linear and tobit models and several alternative one-part and two-part FRMs. In Section 4, we briefly describe some specification tests suited to assessing the conditional mean assumption made in FRMs. In Section 5, we use an empirical example to compare and discuss the different estimators and tests analyzed throughout the paper. In section 6, we present concluding remarks. Throughout the paper, as explained in Section 2, DEA scores are treated as descriptive measures of the relative efficiency of the

DMUs in the sample.

2 Two-stage DEA regression analyses

There are a number of ways in which environmental variables can be accommodated into DEA analysis. Since its introduction by Ray (1991), the standard approach for studying the influence of environmental factors on the efficiency of DMUs is to use two-stage DEA analysis. This involves first using DEA techniques to evaluate the relative efficiency of DMUs and then regressing DEA efficiency scores on appropriate covariates. See Simar and Wilson (2007) for an extensive list of references to the use of this approach, Coelli, Rao, O'Donnell and Battese (2005) and Daraio and Simar (2005) for other approaches to incorporating the influence of efficiency factors into DEA analysis, and Wang and Schmidt (2002) for an explanation of why, in the framework of stochastic frontier analysis, using a single-stage procedure to estimate inefficiency and the impact of environmental variables jointly is the appropriate approach for productivity analysis.

Despite the popularity of two-stage DEA analysis, there has recently been some controversy over applying this approach to examining how a set of environmental factors determines technical efficiency; see Grosskopf (1996) for an early criticism of this approach. Indeed, as pointed out by Simar and Wilson (2007), none of the many studies cited in their paper describes the DGP underlying their two-stage approaches. In order to provide a rationale for second-stage DEA regressions, two distinct justifications have recently been put forward. These justifications depend crucially on the interpretation given to the DEA score used as the dependent variable in the second-stage regression analysis. As discussed by McDonald (2009, Section 11), at stage two, DEA scores may be interpreted either as *observed* measures of DMU efficiency or as *estimates* of 'true', but unobserved, efficiency scores.

In the first of these approaches, which is adopted by, in McDonald's (2009) terminology, 'instrumentalists', DEA scores are treated as descriptive measures of the relative technical efficiency of the sampled DMUs. Given this interpretation, the frontier can be viewed as a (within-sample) observed best-practice construct and, therefore, in stage two, the DEA scores can be treated like any other dependent variable in regression analysis. Hence, parameter estimation and inference in the second stage may be carried out using standard procedures. In this framework, the main issue is choosing an appropriate DGP for the DEA scores, which requires essentially selecting a suitable functional form for the regression model that relates these scores to the environmental variables. In most empirical studies, linear specifications or, because of the bounded nature of DEA scores, tobit models are used. However, as stressed by Simar and Wilson (2007), no coherent account of why a model suitable for censored data should be used in this setting has been provided.

Although the instrumentalist approach may appear simplistic and naive, it reflects the common practice in economics of dealing with dependent variables that are based on the extremely complex measurements of economic aggregates. As McDonald (2009) argues, although the values of such dependent variables should be treated as estimates rather than as actual measures, this is not normally done because 'it is thought it would lead to considerable complexity and perhaps only minor changes in inference'. See McDonald (2009) for more arguments that justify the instrumentalist approach, which has been implicitly adopted for most two-stage DEA empirical studies.

By contrast, in the so-called 'conventionalist' approach (McDonald 2009), DEA scores measure efficiency relative to an estimated frontier (the true value of which is unobserved). This implies that estimates of efficiency from DEA models are subject to uncertainty because of sampling variation. As shown by Kneip, Park and Simar (1998), although DEA scores are consistent estimators of true efficiency, they converge slowly. Moreover, they are biased downwards. Using DEA scores in the second-stage regression analysis leads to two additional problems: (i) the input and output factors used to estimate the DEA scores may be correlated with the explanatory variables in the second stage; and (ii) under the DEA methodology, DEA scores are dependent on each other and, hence, making the assumption of within-sample independence, which is required for regression analysis, is inappropriate. Consequently, the estimated effects of the environmental variables on DMU efficiency may be inconsistent and standard inferential approaches are not valid.

Under the conventionalist framework, a coherent DGP for the DEA scores must include not only a specification for the regression model used in the second stage but also a description of how the variables used in the first and second stage are related. In this context, Simar and Wilson (2007) and Banker and Natarajan (2008) were the first to describe a coherent DGP and to develop appropriate estimation procedures for two-stage

DEA analysis.

Simar and Wilson (2007) provide a set of assumptions under which the consistency of the second-stage regression parameters is not affected by the use of estimates rather than true efficiency scores. To make inferences about those parameters, they propose two alternative bootstrap methods that take into account the sampling variability of DEA scores. Two of the assumptions made by Simar and Wilson (2007) are particularly relevant. First, they assume that a separability condition, which allows environmental variables to affect the efficiency scores but not the frontier, holds. Second, they assume that the true efficiency scores follow a truncated normal distribution. See *inter alia* Zelenyuk and Zheka (2006), Latruffe, Davidova and Balcombe (2008) and Kravtsova (2008) for recent applications of this approach.

Banker and Natarajan (2008) provide a formal statistical foundation for two-stage DEA analyses, deriving conditions under which two-stage procedures yield consistent estimators at stage two. One of their specifications implies a linear relationship between the log of efficiency scores and the environmental variables. This has the favorable implication that the parameters of interest in second-stage DEA analysis can be estimated consistently by using ordinary least squares. The DGP proposed by Banker and Natarajan (2008) is less restrictive than that suggested by Simar and Wilson (2007) (see endnote 1 in the former paper for details). However, they considered only parameter estimation, do not discussing how to testing hypotheses about the parameters estimated in stage two. Moreover, because the dependent variable in the regression model is the log (rather than the level) of the DEA score, reestimating efficiency scores or quantifying the marginal effects requires distributional assumptions about the error term of the second-stage regression.[1]

From our discussion, it is clear that one important issue in both the instrumentalist and conventionalist approaches is the choice of model used in the regression stage. If the specification is incorrect, any of the procedures discussed above will produce inconsistent estimates of the parameters of interest. However, to our knowledge, no one has tested the suitability of the regression model used in second-stage DEA analysis. In this paper, we propose several alternative regression models of efficiency scores (defined on $]0,1]$)

[1] See Duan (1983) for a seminal paper on the consequences for prediction of using logged dependent variables.

on environmental variables, and we show how such specifications may be assessed using simple statistical tests. For simplicity, we adopt the instrumentalist approach and treat DEA scores as observed measures of technical efficiency. This strategy allow us to focus exclusively on two key points the paper: (i) any sensible DGP for DEA scores requires the use of FRMs (or their two-part extensions); and (ii) because there are many alternative FRM specifications, which may generate different results, it is fundamental to test the specification chosen for the regression model.

Thus, while papers such as Simar and Wilson (2007) focus on incorporating into the second-stage the variability induced by estimation in the first stage, and assume that the parametric model of DEA scores is correctly specified, we examine the effects of misspecifying this model in a framework in which such variability is assumed to be absent. Note that misspecification of the functional form is a more serious problem than ignoring the sampling variability of DEA scores: only the former is expected to cause inconsistent estimation of the second-stage parameters of interest. In the final section of this paper, we briefly discuss the implications of using our methodology under the conventionalist approach.

3 Alternative regression models for efficiency scores

Consider a random sample of $i = 1, ..., N$ DMUs. Let y be the variable of interest (the DEA efficiency score), $0 < y \leq 1$, and let x be a vector of k environmental factors. Let $f(y|x, \theta)$ denote the conditional distribution of y, which may be known or unknown, and let θ be the vector of parameters to be estimated. In this section, we first discuss the main characteristics of the traditional regression models employed in the second stage of DEA efficiency analysis. Then, we present the (one- and two-part) FRMs.

3.1 Traditional approaches: linear and tobit models

Some DEA analyses have used linear conditional mean models, given by

$$E(y|x) = x\theta, \tag{1}$$

to explain efficiency scores. However, the linearity assumption is unlikely to hold in the DEA framework, for two main reasons. First, the conceptual requirement that the predicted values of y lie in the interval $]0,1]$ is not satisfied. Second, in a linear model, the marginal effect on the DEA score of a unitary change in covariate x_j,

$$\frac{\partial E(y|x)}{\partial x_j} = \theta_j,$$

is constant over the entire range of y, which is not compatible with either the bounded nature of DEA scores or the existence of a mass point at unity in their distribution.

The traditional approach to explaining DEA scores is to use a two-limit tobit model on data censored at 0 and 1; see the references in Simar and Wilson (2007). This model assumes that there is a latent variable of interest, y^*, $-\infty < y^* < +\infty$, which is not fully observed. Instead of observing y^*, we observe y, which is defined as follows: $y = 0$ if $y^* \leq 0$, $y = y^*$ if $0 < y^* < 1$, and $y = 1$ if $y^* \geq 1$. It is also assumed that there exists a linear relationship between y^* and the covariates, $E(y^*|x) = x\theta$, which implies that the conditional mean of the observed fractional response variable is given by

$$E(y|x) = \left[\Phi\left(\frac{1-x\theta}{\sigma}\right) - \Phi\left(-\frac{x\theta}{\sigma}\right)\right]x\theta - \sigma\left[\phi\left(\frac{1-x\theta}{\sigma}\right) - \phi\left(-\frac{x\theta}{\sigma}\right)\right] + 1 - \Phi\left(\frac{1-x\theta}{\sigma}\right). \tag{2}$$

The partial effects of a unitary change in x_j on y are given by

$$\frac{\partial E(y|x)}{\partial x_j} = \theta_j\left[\Phi\left(\frac{1-x\theta}{\sigma}\right) - \Phi\left(-\frac{x\theta}{\sigma}\right)\right], \tag{3}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal distribution and density functions, respectively, and σ is the standard deviation of the error term $u = y^* - E(y^*|x)$; see *inter alia* Hoff (2007).

Researchers such as Simar and Wilson (2007) and McDonald (2009) criticize this approach on the grounds that the concentration of observed DEA scores at unity is a product of the way the scores are defined rather than the result of a censoring mechanism, as implied by the tobit model. Indeed, whereas with censored data one is typically interested in inferring the effects of a change in x_j on y^*, which is given simply by θ_j, in the DEA framework, one is interested in the partial effects on the observable variable y, given in (3); see Wooldridge (2002) pp. 517-521 for discussion of these issues. However, some DEA researchers mistakenly focus on the potential marginal effects, given by θ_j, rather than on the actual partial effects (3); see, e.g., Chilingerian (1995) and Gillespie, Schupp and Taylor (1997).

As pointed out by Wooldridge (2002), provided that the focus of the research is changed from y^* to y, tobit and other regression models originally devised for censored dependent variables may also be used for response variables that are by nature limited from above and/or below (so-called 'corner solution' variables). According to this view, the tobit model described by equation (2) is a plausible specification for the conditional mean of a variable defined on the interval $[0, 1]$. However, in the DEA framework, efficiency scores do not generally take on values of zero. In the absence of observations for $y = 0$, the first term of the log-likelihood function of the two-limit tobit model,

$$LL_i = I_{(y_i=0)} \log \Phi \left(-\frac{x_i\theta}{\sigma} \right) + I_{(0<y_i<1)} \log \frac{1}{\sigma} \phi \left(\frac{y_i - x_i\theta}{\sigma} \right) + I_{(y_i=1)} \log \left[1 - \Phi \left(\frac{1 - x_i\theta}{\sigma} \right) \right], \tag{4}$$

disappears. This means that, in practice, estimation is based on a one-limit tobit model for $y \in]-\infty, 1]$. Although the consequences of this are not usually serious, it is obvious that the DGP that underpins this tobit model is not the one that governs the variable of interest.

3.2 Fractional regression models

A model that avoids the problems associated with using linear and tobit models in the DEA framework is the FRM. This model was proposed by Papke and Wooldridge (1996) to deal with dependent variables defined on the unit interval, irrespective of whether boundary values are observed. In fact, given that DEA scores are relative measures of efficiency, they can be seen as the result of a normalizing DGP in which the efficiency measures are mapped onto the interval $]0, 1]$; see McDonald (2009).

The FRM only requires the assumption of a functional form for y that imposes the desired constraints on the conditional mean of the dependent variable, as follows:

$$E(y|x) = G(x\theta), \tag{5}$$

where $G(\cdot)$ is some nonlinear function satisfying $0 \leq G(\cdot) \leq 1$. The model defined by (5) may be consistently estimated by QML as suggested by Papke and Wooldridge (1996). Alternatively, nonlinear least squares or maximum likelihood estimation may be used, but the former is less efficient than QML estimation, and the latter requires the specification of the conditional distribution of y given x (for which the beta distribution is commonly chosen).

Papke and Wooldridge (1996) propose estimating FRMs by QML based on the Bernoulli log-likelihood function, which is given by

$$LL_i(\theta) = y_i \log[G(x_i\theta)] + (1 - y_i)\log[1 - G(x_i\theta)]. \tag{6}$$

Given that the Bernoulli distribution is a member of the linear exponential family, the QML estimator of θ, defined by

$$\hat{\theta} \equiv \arg\max_{\theta} \sum_{i=1}^{N} LL_i(\theta), \tag{7}$$

is consistent and asymptotically normal, regardless of the true distribution of y conditional on x, provided that $E(y|x)$ in (5) is indeed correctly specified (see Gourieroux, Monfort and Trognon 1984 for details). Moreover, as Papke and Wooldridge (1996) point out, there are cases in which this QML estimator is efficient within the class of estimators containing all linear exponential family-based QML and weighted nonlinear least squares estimators. The asymptotic distribution of the QML estimator is given by

$$\sqrt{N}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V), \tag{8}$$

where $V = A^{-1}BA^{-1}$, with $A = E[-\nabla_{\theta\theta'}LL(\theta)]$ and $B = E[\nabla_{\theta}LL(\theta)\nabla_{\theta'}LL(\theta)]$. Consistent estimators for A and B are given by $\hat{A} = N^{-1}\sum_{i=1}^{N}\hat{g}_i^2 x_i' x_i\left[\hat{G}_i\left(1 - \hat{G}_i\right)\right]^{-1}$ and $\hat{B} = N^{-1}\sum_{i=1}^{N}\hat{u}_i^2\hat{g}_i^2 x_i' x_i\left[\hat{G}_i\left(1 - \hat{G}_i\right)\right]^{-2}$, respectively, where $\hat{G}_i \equiv G\left(x_i\hat{\theta}\right)$, $g(x_i\theta) = \partial G(x_i\theta)/\partial(x_i\theta)$, $\hat{g}_i \equiv g\left(x_i\hat{\theta}\right)$ and $\hat{u}_i = y_i - \hat{G}_i$.

3.2.1 Standard models

Papke and Wooldridge (1996) suggest as possible specifications for $G(\cdot)$ any cumulative distribution function, such as those commonly used to model binary data. The most widely used functions are undoubtedly the logit and probit functional forms, where $G(x\theta) = e^{x\theta}/\left(1 + e^{x\theta}\right)$ and $G(x\theta) = \Phi(x\theta)$, respectively. However, there exist alternatives such as the loglog and complementary loglog (hereafter cloglog) specifications, where $G(x\theta) = e^{e^{-x\theta}}$ and $G(x\theta) = 1 - e^{-e^{x\theta}}$, respectively. In all cases the partial effects are given by

$$\frac{\partial E(y|x)}{\partial x_j} = \theta_j g(x\theta). \tag{9}$$

Figure 1 illustrates the four alternative functional forms for $G(x\theta)$ referred to, as well as the corresponding $g(x\theta)$ functions that appear in (9). While the symmetric logit and probit models approach zero and unity at the same rate, the asymmetric cloglog (loglog) model increases slowly (sharply) at small values of $G(\cdot)$ and sharply (slowly) when $G(\cdot)$ is near unity. On the other hand, the maximum partial effects produced by the symmetric models are achieved at $E(y|x) = 0.5$ and are identical for values of x that yield values of $E(y|x)$ that are symmetric around that point: e.g., the effect of x_j on $E(y|x)$ is the same for $E(y|x) = 0.05$ and $E(y|x) = 0.95$. By contrast, in the cloglog (loglog) model, the greatest impact of a change in x_j occurs on DMUs with $E(y|x) > 0.5$ ($E(y|x) < 0.5$).

Figure 1 about here

3.2.2 Generalized models

The models analyzed in the previous section impose *a priori* the condition that DMUs with a given efficiency score (e.g., 0.5 in symmetric models), say y_a, are the most sensitive to changes in the explanatory variables. However, if DMUs with efficiency scores other than y_a are the ones that are most sensitive to such changes, then the assumed model is misspecified and leads to biased inferences about the marginal effect of any independent variable. In this section, we discuss some alternative FRMs in which y_a is not set *a priori* but is instead determined by the actual patterns observed in the data.

The new models that we propose for DEA regressions are based on two alternative generalizations of the specifications analyzed in the previous section, both of which use an additional parameter, α, to modify the form of the response curves illustrated in Figure 1. The first generalization (generalized type I model) consists simply of raising to α any functional form $G(x\theta)$ appropriate for dealing with fractional response variables,

$$E(y|x) = G(x\theta)^{\alpha}. \tag{10}$$

The second generalization, based on the asymmetry of a complementary form (generalized type II model), is given by

$$E(y|x) = 1 - [1 - G(x\theta)]^{\alpha}, \tag{11}$$

where $\alpha > 0$ such that $0 < E(y|x) < 1$. Both generalizations reduce to $G(x\theta)$ for $\alpha = 1$ and have already been considered by Ramalho, Ramalho and Murteira (2010) to derive score tests for the adequacy of the nested functional form $G(x\theta)$. However, these regression models have not been used to analyze fractional response variables. For binary logit models, similar (but not identical) generalizations have already been considered; see *inter* *alia* Poirier (1980), Smith (1989) and Nagler (1994). For other generalizations commonly employed with binary models, which can, however, only be applied to specific functional forms, usually the logit, see *inter alia* Prentice (1976), Pregibon (1980), Aranda-Ordaz (1981), Whitemore (1983), Stukel (1988) and Czado (1994).

The generalized models for $E(y|x)$ in (10) and (11) describe a wide variety of asymmetric patterns, giving rise to flexible regression models. The forms of asymmetry created by the introduction of α into the functional form of $E(y|x)$ are illustrated in Figure 2 (first column) for logit models (similar patterns would be obtained for other models). Clearly, the value of α determines the magnitude and the direction of the shift in the standard logit curve. In particular, for both generalizations, the magnitude of the shift is larger the farther is α from unity, with the generalization (10) ((11)) shifting the original logit curve to the right for $\alpha > 1$ ($0 < \alpha < 1$) and to the left for $0 < \alpha < 1$ ($\alpha > 1$), with a more substantial impact on the left (right) tail.

Figure 2 about here

The partial effects of a unitary change in x_j in models (10) and (11) are now given by

$$\frac{\partial E(y|x)}{\partial x_j} = \theta_j g(x\theta)\,\alpha G(x\theta)^{\alpha-1} \tag{12}$$

and

$$\frac{\partial E(y|x)}{\partial x_j} = \theta_j g(x\theta)\,\alpha\left[1 - G(x\theta)\right]^{\alpha-1}, \tag{13}$$

respectively. These partial effects are illustrated in Figure 2 (second column) for the logit case. Again, it is clear that α governs both the magnitude and the asymmetric shape of the curves of the partial effects. It is also clear that an infinite variety of asymmetric shapes can be generated. Moreover, the value of y_a also depends on α, which implies that the greatest impact of a change in x_j is allowed to occur for any efficiency score. In model (10) ((11)), when more efficient DMUs are more sensitive to changes in x_j, the value of α should be high (low); when less efficient DMUs are more sensitive to changes in x_j, the value of α should be low (high).

Despite their clear advantages in terms of flexibility, the two generalized FRMs we propose should be used with care in applied work, particularly when estimation is based on small samples. Indeed, the addition of the extra parameter α is expected to increase substantially the variance of the estimates in many cases. For example, Taylor (1988) analyzed the two generalized binary-response regression models proposed by Aranda-Ordaz (1981), which are also based on the incorporation of an additional parameter. Using Monte Carlo methods, Taylor (1988) found that the cost of adding that parameter in terms of variance inflation is 50% on average, although it can be appreciably larger or smaller. In our empirical application, we sometimes encountered a similar problem. Thus, particularly with small samples, we recommend using the proposed generalizations only when all the standard specifications discussed in the previous section prove to be inadequate.

3.3 Two-part models

All models discussed so far assume that the same environmental variables affect efficient and inefficient DMUs in the same way. However, when the probability of observing a DEA score of unity is relatively large, one may suspect that the sources of DMU efficiency may differ from those of DEA inefficiency. For example, a given environmental factor may have a significant effect on the probability of observing an efficient DMU ($y = 1$) but may not explain the degree of inefficiency of DMUs for which $y < 1$.

In such a case, a two-part model should be used for modeling DEA scores. The first part of such a model comprises a standard binary choice model that governs the probability of observing an efficient DMU. Let z be a binary indicator that takes the values of zero and unity for inefficient and efficient DMUs, respectively, as follows:

$$z = \begin{cases} 1 \text{ for } y = 1 \\ 0 \text{ for } 0 < y < 1. \end{cases} \tag{14}$$

Assume also that the conditional probability of observing an efficient DMU is

$$\Pr(z = 1|x) = E(z|x) = F(x\beta_{1P}), \tag{15}$$

where β_{1P} is a vector of variable coefficients and $F(\cdot)$ is a cumulative distribution function, typically one of those discussed in section 3.2.1 for FRMs, although the generalized specifications in (10) and (11) may also be used. Equation (15) may be estimated by maximum likelihood using the whole sample.

The second part of the model is estimated using only the sub-sample of inefficient DMUs and governs the magnitude of the DEA scores on the interval $]0,1[$:

$$E(y|x, y \in]0,1[) = M(x\beta_{2P}), \tag{16}$$

where $M(\cdot)$ may be any of the specifications considered for $E(y|x)$ in the previous section and β_{2P} is another vector of coefficients.

The partial effects of a covariate x_j over the probability of observing an efficient DMU and the conditional mean DEA score for an inefficient DMU are given by, respectively,

$$\frac{\partial \Pr(z=1|x)}{\partial x_j} = \frac{\partial F(x\beta_{1P})}{\partial x_j} = \beta_{1P} f(x\beta_{1P}) \tag{17}$$

and

$$\frac{\partial E(y|x, y \in]0,1[)}{\partial x_j} = \frac{\partial M(x\beta_{2P})}{\partial x_j} = \beta_{2P} m(x\beta_{2P}), \tag{18}$$

where $f(x\beta_{1P})$ and $m(x\beta_{2P})$ are the partial derivatives of $F(\cdot)$ and $M(\cdot)$ with respect to $x\beta_{1P}$ and $x\beta_{2P}$, respectively. For simplicity, we assume that the same regressors appear in both parts of the model, but this assumption can relaxed and, in fact, should be if there are obvious exclusion restrictions.

The overall conditional mean and the partial effects of x_j on y can be written as

$$\begin{aligned} E(y|x) &= E(y|x, y \in]0,1[) \cdot \Pr[y \in]0,1[\,|x] + E(y|x, y=1) \cdot \Pr(y=1|x) \\ &= E(y|x, y \in]0,1[) \cdot \Pr[y \in]0,1[\,|x] + \Pr(y=1|x) \\ &= M(x\beta_{2P}) \cdot [1 - F(x\beta_{1P})] + F(x\beta_{1P}) \end{aligned} \tag{19}$$

and

$$\begin{aligned} \frac{\partial E(y|x)}{\partial x_j} &= \frac{\partial M(x\beta_{2P})}{\partial x_j}[1 - F(x\beta_{1P})] - M(x\beta_{2P})\frac{\partial F(x\beta_{1P})}{\partial x_j} + \frac{\partial F(x\beta_{1P})}{\partial x_j} \\ &= \beta_{2P} m(x\beta_{2P})[1 - F(x\beta_{1P})] + \beta_{1P} f(x\beta_{1P})[1 - M(x\beta_{2P})]. \end{aligned} \tag{20}$$

Thus, the total change in y can be disaggregated in two parts: (i) the change in the DEA scores of inefficient DMUs, weighted by the probability of observing such DMUs; and (ii) the change in the probability of observing an efficient DMU, weighted by one minus the expected efficiency score of an inefficient DMU. This decomposition is similar to that used by McDonald and Moffitt (1980) for the tobit model but does not constrain β_{1P} and β_{2P} to be identical or require $F(\cdot)$ and $M(\cdot)$ to be based on normal distribution functions.

To illustrate the rich variety of partial effects that may be produced by two-part models, we consider in Figure 3 some specific cases. Given that the partial effects (20) may be expressed as a function of both $\Pr(y=1|x)$ and $E(y|x, y \in]0,1[)$, in Figure 3, we set the former at 0.2 or 0.8 and examine how the partial effects change as the latter increases from zero to unity. Of the many potential two-part models that can be constructed, in Figure 3, we consider only eight variants, all of which use a logit model in the first or second part of the model. As a reference, we also consider the one-part logit model. Figure 3 clearly illustrates that one can obtain results that differ substantially from those from the simple logit model.

Figure 3 about here

The two-part model considered by Hoff (2007), termed the unit inflated beta model, is much more restrictive than the one we propose. Indeed, Hoff (2007) assumes a logit specification for $\Pr(z = 1|x)$ in (15), whereas we allow for many other alternative specifications, the adequacy of each of which is easily tested. Moreover, Hoff (2007) also uses a logit specification for $E(y|x, y \in]0, 1[)$ and assumes a beta distribution for the DEA scores of inefficient DMUs. By contrast, in our model, we need only specify the conditional mean DEA score, for which we also consider many alternative (and easily testable) functional forms.

4 Specification tests

As our exposition suggests, the main practical difference between the alternative one-part and two-part regression models discussed in the previous section relates to the functional form assumed for $E(y|x)$: see (5), (10), (11) and (19). In fact, for any of those regression models, correctly formalizing $E(y|x)$ is critical for consistent estimation of the parameters of interest. However, despite the availability of a number of tests that can be used for testing conditional mean assumptions, such tests have not been applied in empirical studies on second-stage DEA.

The correct specification of the functional form of the conditional mean $E(y|x)$ requires that a correct model must be specified for $G(x\theta)$ and for both $F(x\beta_{1P})$ and $M(x\beta_{2P})$ in the one- and two-part models, respectively. Moreover, the data must be governed by the one-part or two-part mechanisms assumed. The reason for this is that even if the covariates of the first and second parts of the model coincide, in general, one-part models are not appropriate for data governed by two-part models, and vice versa. Therefore, in this section, we discuss general tests to assess these two aspects.

One way of assessing whether the specification of $E(y|x)$ is correct is to use tests appropriate for detecting general functional form misspecifications, such as the well-known

RESET test. Indeed, using standard approximation results for polynomials, it can be shown that any index model of the form $E(y|x) = L(x\theta)$ can be arbitrarily approximated by $S\left(x\theta + \sum_{j=1}^{J} \gamma_j (x\theta)^{j+1}\right)$ if J is sufficiently large; see Pagan and Vella (1989). Therefore, testing the hypothesis $E(y|x) = S(x\theta)$ is equivalent to testing $\gamma = 0$ in the augmented model $E(y|x,z) = S(x\theta + z\gamma)$, where $z = \left[\left(x\hat{\theta}\right)^2, ..., \left(x\hat{\theta}\right)^{J+1}\right]$. The first few terms in the expansion are the most important, and, in practice, only the quadratic and cubic terms are usually considered. Note that the RESET test cannot be directly applied to assess (19), the functional form assumed for two-part models. Instead, it has to be separately applied to their two components, given by (15) and (16).

Alternatively, because all competing specifications for $E(y|x)$ are non-nested, we may apply standard tests for non-nested hypotheses, where the alternative competing specifications for $E(y|x)$ are tested against each other. An example of this type of test is the P test proposed by Davidson and MacKinnon (1981), which is probably the simplest way of comparing nonlinear regression models; see *inter alia* Gourieroux and Monfort (1994) for other alternatives. To our knowledge, the P test has not been applied either in empirical DEA studies or for choosing between linear, tobit and FRMs; however, Ramalho, Ramalho and Murteira (2010) use the P test for discriminating between alternative one-part and two-part FRMs.

Suppose that $H(x\alpha)$ and $T(x\eta)$ are competing functional forms for $E(y|x)$. As shown by Davidson and MacKinnon (1981), testing $H_0 : H(x\alpha)$ against $H_1 : T(x\eta)$ (i.e., checking whether $H(x\alpha)$ is an appropriate specification for $E(y|x)$ after taking into account the information provided by the alternative model) is equivalent to testing the null hypothesis $H_0 : \delta_2 = 0$ in the following auxiliary regression:

$$\left(y - \hat{H}\right) = \hat{h}x\delta_1 + \delta_2\left(\hat{T} - \hat{H}\right) + error, \tag{21}$$

where $h = \partial H(x\alpha) / \partial (x\alpha)$, δ_2 is a scalar parameter and $\hat{}$ denotes evaluation at the estimators $\hat{\alpha}$ or $\hat{\eta}$, obtained by separately estimating the models defined by $H(\cdot)$ and $T(\cdot)$, respectively. To test $H_0 : T(x\eta)$ against $H_1 : H(x\theta)$, we need to use another P statistic, which is calculated using a similar auxiliary regression to (21) but with the roles of the two models interchanged. As is standard with tests of non-nested hypotheses, three outcomes are possible: one may reject one model and accept the other, accept both models or reject both.

The P test based on (21) may be used for choosing between: (i) various possible specifications for one-part models, i.e., those given in (1), (2), (5), (10) and (11); (ii) one-part and two-part models, i.e., (1), (2), (5), (10) or (11) versus (19); and (iii) alternative specifications for two-part models, i.e., (19). In addition, $H(x\alpha)$ and $T(x\eta)$ may represent alternative functional forms for $\Pr(z=1|x)$ or $E(y|x, y \in [0,1[)$, in which case, the P test may be used to select between competing specifications for the first or the second component of a two-part model, respectively.

One may also apply the GOFF-I and GOFF-II tests proposed by Ramalho, Ramalho and Murteira (2010) to determine, respectively, whether the Type I or Type II generalizations are indeed necessary or, instead, the corresponding simpler standard FRM is adequate. Moreover, as the conditional variance of y is in general a function of its conditional mean, because the former must change as the latter approaches either boundary, heteroskedasticity-robust versions must be computed in all cases. In the empirical application that follows, we compute robust LM versions of all tests, which have the advantage of not requiring the estimation of any alternative model; see Papke and Wooldridge (1996) and Ramalho, Ramalho and Murteira (2010) for details of the computation of those statistics. Robust Wald statistics may also be easily obtained using statistical packages such as Stata.

5 Empirical application

In this section, we apply the techniques described so far to the regression analysis of DEA efficiency scores of Portuguese farms. First, we provide a brief description of the data used in the analysis. Then, we illustrate the usefulness of the specification tests discussed in Section 4 for selecting appropriate regression models for DEA scores. Finally, we compare the regression results of linear, tobit and some FRMs in the following respects: (i) the significance and sign of the estimated parameters of interest; (ii) the magnitude of the partial effects; and (iii) the prediction of DEA efficiency scores.

5.1 Data

The data set used in this study is based on individual farm account records collected by the Portuguese Ministry of Agriculture for the year 2004. Our sample comprises a total of 266 farms located in the Portuguese region of Alentejo. In the first stage of DEA analysis, we considered a single output, performed an input-oriented study and assumed variable returns to scale and weak free disposability of inputs. For each farm, output was measured as the gross value of production and for inputs we considered two classes of capital, two categories of labor and two categories of land.

According to this DEA analysis, 117 (44%) of the sampled farms produced on the efficiency frontier, which makes this case study particularly relevant for comparing the alternative second-stage DEA models discussed in this paper. The average DEA score is 0.781, and the quartiles of its distribution are 0.585, 0.867 and 1. The lowest score is 0.213.

In the illustrative second stage of DEA analysis that follows, the efficiency scores were related to the following factors: land ownership, farm specialization, economic size, farm subsidies and geographical location. Land ownership is represented by a dummy variable, LANDLORD, which takes the value of unity if the farmer owns the land and 0 otherwise. Farm specialization is represented by two dummy variables, LIVESTOCK and CROP, which take the value of unity if the farm specializes in livestock or crops, respectively, and zero otherwise. Size (SIZE) is measured as the farm's volume of sales. Farm subsidies (SUBSIDIES) are measured as the proportion of subsidies in the farm's total revenue. In addition, because Alentejo is usually divided into four Nut III regions - Alto Alentejo, Alentejo Central, Baixo Alentejo and Alentejo Litoral - we used the dummy variables ALTO, CENTRAL and BAIXO, which take the value of unity if the farm is located in the corresponding region and zero otherwise. These factors were used as explanatory variables in all regression models, including the two components of the two-part models.

5.2 Model selection

For the second stage of the DEA analysis, we have available a large set of alternative specifications: traditional models (linear and tobit), one-part standard and generalized FRMs, which may be based on logit, probit, loglog or cloglog functional forms, and two-part models that, in each part, use any of the previous (standard or generalized) functional forms (plus a linear one in the second part). Given the large set of models that can be estimated, for the one-part models and for the two components of the two-part models, we start our empirical analysis by applying to each alternative formalization the following tests: the RESET test (based on one fitted power of the response index); the P test, considering, one by one, all the other possible specifications as the alternative hypothesis; and, when applicable, Ramalho, Ramalho and Murteira's (2010) GOFF-I and GOFF-II tests.

Table 1 summarizes the results obtained for the one-part models. These results clearly indicate that only a few specifications are admissible. Indeed, only the cloglog model and its type I generalization are never rejected at the 10% level. Moreover, almost all of the other specifications are rejected when the P test uses one of the acceptable models as the alternative hypothesis. Given that the GOFF-I test does not reject the correct specification of the standard cloglog model and given that the size of our sample is relatively small, we select the cloglog model as the most suitable one-part model. In fact, given that the distribution of efficiency scores in our example is clearly asymmetric and given that the number of one outcomes is large, a cloglog functional form would be our preferred choice of one-part model.

Table 1 about here

As shown in Table 2, we choose also a cloglog specification for explaining the probability of a farm producing on the efficiency frontier. In contrast, as Table 3 shows, all tests fail to reject any of the 13 models estimated for the second component of the two-part models, including the linear model. This suggests that the main issue in the regression analysis of DEA scores is not so much their bounded nature as the existence of a mass-point at unity in their distribution. Because the GOFF tests do not reject any of the simpler functional forms, we proceed by considering five alternative two-part models, which use a cloglog specification in the first part and a linear, logit, probit, loglog or cloglog model in the second part.

Table 2 about here

Table 3 about here

We also applied versions of the P test that enable testing of the selected cloglog one-part model against the full specification of the five selected two-part models, and vice versa, and testing of the selected full specification of the two-part models, with each being tested against the others. In no case were we able to reject any of the tested specifications.[2]

[2] Full results are available from the authors on request.

5.3 Regression results

In Table 4, we report the estimation results obtained from the selected models and compare them with those from the linear, tobit and the most commonly used FRM, the logit model. For each explanatory variable, we report the value of the associated estimated coefficient and its standard error. For each model, we report also the percentage of predictions outside the unit interval and the R^2, which was calculated as the square of the correlation between the actual and predicted efficiency scores and, thus, is comparable across models and over estimation methods.

Table 4 about here

Although most of the R^2 values are similar, they provide further evidence that the selected models fit the data at least as well as the competing models. Indeed, the highest R^2s among the one-part models and the first component of the two-part models are for the selected cloglog models. On the other hand, the R^2s of the alternative specifications considered for the second stage of the two-part models are virtually identical, which further confirms that when farms on the frontier are excluded from the regression analysis, most functional forms are in general adequate for modeling DEA scores. However, note that even in this case, for a few cases, the linear model yields predicted outcomes that exceed unity.

The first striking point to emerge from the analysis of the regression coefficients displayed in Table 4 is that while all estimators produce the same conclusions in terms of their sign and significance in the two-part models, the same does not happen with one-part models. Indeed, in the latter case, there are some explanatory variables that have significant coefficients in some models but not in others. Note in particular the clear differences between the tobit and the selected cloglog models: there are three (one) coefficients that are significant at the 10% level in the tobit (cloglog) model but not in the cloglog (tobit). Again, these differences seem to be a consequence of the difficulty that most one-part models have in dealing with a large proportion of DMUs taking the value of unity.

Another interesting point is that the number of significant coefficients is much larger in the two-part models. That is, analyzing separately, first, why some farms are on the efficiency frontier and others are not, and, second, the distance to the frontier of the inefficient farms, seems to be a better way of uncovering the real effect of each covariate on the DEA score. An example is the case of the variable LIVESTOCK. According to the one-part models, specializing in livestock does not significantly affect a farm's efficiency. However, the two-part models show clearly that farms specializing in livestock or crops are more likely to be on the efficiency frontier. On the other hand, in the sub-sample of only inefficient farms, specializing in livestock lowers the DEA score.

In Table 5, we report for each model the partial effects estimated for each covariate, which were calculated as the mean of the partial effects computed for each farm in the sample. These results confirm that the functional form chosen for the second stage of the two-part models hardly affects the results. By contrast, choosing the wrong model for the first stage may seriously bias the estimation of the partial effects. For example, if instead of using the cloglog model selected by the specification tests, we had decided to use the commonly adopted logit model, the bias is over 10% for several covariates (LANDLORD, SIZE, SUBSIDIES and ALTO), assuming that the cloglog is indeed the correct model. On the other hand, in the one-part models, the differences between the various alternative specifications may be substantial. For example, among the linear, tobit and logit regression coefficients, the maximum differences relative to the coefficients of the selected cloglog model are 42% (SIZE), 52% (LIVESTOCK) and 41% (SUBSIDIES), respectively. A comparison of the partial effects implied by the one- and two-part models suggests that even the models selected by the specification tests may generate very different results.

Table 5 about here

In Figure 4, for specific models, we report partial effects and predicted DEA scores as a function of SIZE, with the other covariates set at their mean values (SUBSIDIES) or at their modes (the dummy variables). In this representation, for SIZE, we consider 1000 equally spaced values between the 0.01 and 0.99 quantiles of its sample distribution. The first graph of Figure 4 clearly illustrates that linearity, which is assumed in many existing second-stage DEA analyses and which implies constant partial effects, may provide conclusions that differ substantially from those implied by the models selected by the specification tests (and, in fact, from all the other non-linear specifications considered in this paper). Indeed, all the other models indicate that the effect of SIZE on the farm's efficiency is much larger for smaller farms. Regarding the prediction of DEA scores, which is useful, for example, for computing the extent of managerial inefficiency not caused by external factors (e.g. Ray, 1991), the differences between the various competing models are not major. Note, however, that both the tobit model and particularly the linear model underestimate the efficiency scores for most values of SIZE.

Figure 4 about here

6 Concluding remarks

In this paper, we considered various alternative approaches to second-stage DEA regressions. We argued that the DGP that governs DEA scores is not appropriately represented by linear or tobit regression models, which are the standard approaches to second-stage DEA analysis. We have shown that, instead, using FRMs are the most natural way of modeling bounded, proportional response variables such as DEA scores. Because some DEA scores are unity, we discussed both one- and two-part FRMs. Tests for assessing the correct specification of each alternative model were also reviewed.

In our empirical example, we found that the main issue in the regression analysis of DEA scores is not so much their bounded nature as the existence of a mass point at unity in their distribution. Therefore, two-part models may be useful in this framework, particularly when the percentage of unity values is large. We found important differences between the FRMs selected by the specifications tests and the linear, tobit and other FRMs, particularly in terms of the magnitudes of the partial effects generated by each competing model. Given the variety of FRMs that can be constructed, functional form tests suited to selecting the most adequate model should be routinely applied in second-stage DEA analysis.

In this paper we made the crucial assumption, common in the existing literature, that DEA scores are descriptive measures of the relative performance of DMUs. As discussed in Section 2, particularly since the publication of the seminal paper by Simar and Wilson (2007), some researchers have criticized this approach on the grounds that DEA scores should be viewed as estimates rather than actual observations and that second-stage DEA analyses should take this into account. Therefore, our paper could be usefully extended by applying our proposed estimating and testing procedures to the 'conventionalist' approach. Such an extension would have to deal with two main issues.

First, it must be shown that using estimated DEA scores ($\hat{y}$) rather than observed DEA scores (y) does not generate inconsistent parameter estimation. Given that $\hat{y}$ converges to y, albeit at a slow rate, one might expect consistency to be maintained, provided that an appropriate set of assumptions (such as a separability condition of the type assumed by Simar and Wilson 2007) is made. However, because $\hat{y}$ is a biased estimator of y, in small samples, the estimators produced by our methods for the second-stage regression parameters will also be biased. However, this small-sample bias seems to be a common feature of all the estimators proposed in the DEA literature.

The second issue that any extension of our methodology to the conventionalist approach has to deal with is how to make inferences about the regression parameters. Indeed, the standard errors and test statistics obtained from standard procedures are generally invalid because they ignore the sampling variability in $\hat{y}$. Given that deriving the asymptotic distribution of the estimators of the second-stage regression parameters would be a formidable task (neither Simar and Wilson 2007 nor Banker and Natarajan 2008 did so for their models), bootstrap procedures similar to those proposed by Simar and Wilson (2007) seem to be the only feasible way to make valid inference in this framework. In order to implement (adaptations of) their parametric bootstraps, additional distributional assumptions are required. For example, for one-part models, we may assume that DEA scores have a beta or a simplex distribution with mean given by our (5), (10) or (11), which would replace assumptions A2 and A3 of Simar and Wilson (2007); see Ramalho, Ramalho and Murteira (2010) for details of these distributions. Under these additional assumptions, both of the bootstrap methods suggested by Simar and Wilson (2007), with the necessary adaptations, could be straightforwardly applied in our framework. For two-part models, one could proceed similarly, assuming that one of the distributions suggested above is applied to the DEA scores in the second part of the model. In all cases, given the additional distributional assumptions made, one would have to apply additional specification tests, such as the information matrix tests referred to by Ramalho, Ramalho and Murteira (2010).

References

Aranda-Ordaz, F.J. (1981), "On two families of transformations to additivity for binary response data", *Biometrika*, 68(2), 357-363.

Banker, R.D. and Natarajan, R. (2008), “Evaluating contextual variables affecting productivity using data envelopment analysis”, *Operations Research*, 56(1), 48–58.

Chilingerian, J.A. (1995), “Evaluating physician efficiency in hospitals: a multivariate analysis of best practices”, *European Journal of Operational Research*, 80, 548–574.

Coelli, T.J., Rao, D.S., O’Donnell, C.J. and Battese, G.E. (2005), *An Introduction to Efficiency and Productivity Analysis*, 2nd ed., Springer.

Czado, C. (1994), “Parametric link modification of both tails in binary regression”, *Statistical Papers*, 35, 189-201.

Daraio, C. and Simar, L. (2005), “Introducing environmental variables in nonparametric frontier models: a probabilistic approach”, *Journal of Productivity Analysis*, 24, 93-121.

Davidson, R. and MacKinnon, J.G. (1981), “Several tests for model specification in the presence of alternative hypotheses”, *Econometrica*, 49(3), 781-793.

Duan, N. (1983), "Smearing estimate: a nonparametric retransformation method", *Journal of the American Statistical Association*, 78, 605-610.

Gillespie, J, Schupp, A. and Taylor, G. (1997), “New alternative enterprise: the case of ratite industry”, *Journal of Agricultural and Applied Economics*, 29(2), 409–418.

Gouriéroux, C. and Monfort, A. (1994), “Testing non-nested hypotheses”, in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics*, Volume IV, Elsevier Science, pp. 2585-2637.

Gouriéroux, C., Monfort, A. and Trognon, A. (1984), “Pseudo maximum likelihood methods: applications to Poisson models”, *Econometrica*, 52(3), 701-720.

Grosskopf, S. (1996), “Statistical inference and nonparametric efficiency: a selective survey”, *Journal of Productivity Analysis*, 7, 161-176.

Hoff, A. (2007), “Second stage DEA: comparison of approaches for modelling the DEA score”, *European Journal of Operational Research*, 181, 425–435.

Kneip, A., Park, B.U. and Simar, L. (1998), "A note on the convergence of nonparametric DEA estimators for production efficiency scores", *Econometric Theory*, 14, 783-793.

Kravtsova, V. (2008), "Foreign presence and efficiency in transition economies", *Journal of Productivity Analysis*, 29, 91–102.

Latruffe, L, Davidova, S. and Balcombe, K. (2008), "Application of a double bootstrap to investigation of determinants of technical efficiency of farms in Central Europe", *Journal of Productivity Analysis*, 29, 183-191.

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

McDonald, J.F. and Moffitt, R.A. (1980), "The uses of tobit analysis", *Review of Economics and Statistics*, 62(2), 318-321.

Nagler, J. (1994), "Scobit: an alternative estimator to logit and probit", *American Journal of Political Science*, 38(1), 230-255.

Pagan, A. and Vella, F. (1989), "Diagnostic tests for models based on individual data: a survey", *Journal of Applied Econometrics*, 4, S29-S59.

Papke, L.E. and Wooldridge, J.M. (1996), "Econometric methods for fractional response variables with an application to 401(k) plan participation rates", *Journal of Applied Econometrics*, 11(6), 619-632.

Poirier, D.J. (1980), "A Lagrange multiplier test for skewness in binary models", *Economics Letters*, 5, 141-143.

Pregibon, D. (1980), "Goodness of link tests for generalized linear models", *Applied Statistics*, 29(1), 15-24.

Prentice, R.L. (1976), "A generalization of the probit and logit methods for dose response curves", *Biometrics*, 32, 761-768.

Ray, S.C. (1991), "Resource-use efficiency in public schools: a study of Connecticut data", *Management Science*, 37(12), 1620-1628.

Ramalho, E.A., Ramalho, J.J.S. and Murteira, J. (2010),"Alternative estimating and testing empirical strategies for fractional regression models", *Journal of Economic Surveys*, forthcoming.

Ruggiero, J. (1998), "Non-discretionary inputs in data envelopment analysis", *European Journal of Operational Research*, 111, 461–469.

Simar, L. and Wilson, P. (2007), "Estimation and inference in two-stage, semi-parametric models of production processes", *Journal of Econometrics*, 136, 31-64.

Smith, R.J. (1989), "On the use of distributional misspecification checks in limited dependents variable models", *Economic Journal*, 99, 178-192.

Stukel, T.A. (1988), "Generalized logistic models", *Journal of the American Statistical Association*, 83(402), pp. 426-431.

Taylor, J.M.G. (1988), "The cost of generalizing logistic regression", *Journal of the American Statistical Association*, 83(404), pp. 1078-1083.

Wang, H. and Schmidt, P. (2002), "One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels", *Journal of Productivity Analysis*, 18, 129-144.

Whitemore, A.S. (1983), "Transformations to linearity in binary regression", *SIAM Journal of Applied Mathematics*, 43(4), pp. 703-710.

Wooldridge, J.M. (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.

Zelenyuk, V. and Zheka, V. (2006), "Corporate governance and firm's efficiency: the case of a transitional country, Ukraine", *Journal of Productivity Analysis*, 25, 143-157.

Table 1: Specification tests for one-part models (p-values)

	Linear	Tobit	FRM				GFRM-I				GFRM-II			
			Logit	Probit	Loglog	Cloglog	Logit	Probit	Loglog	Cloglog	Logit	Probit	Loglog	Cloglog
RESET test	0.686	0.065*	0.048**	0.066*	0.045**	0.148	0.693	0.266	0.059*	0.738	0.219	0.025**	0.528	0.143
GOFF-I test	—	—	0.043**	0.070*	—	0.149	—	—	—	—	—	—	—	—
GOFF-II test	—	—	0.050**	0.062*	0.033**	—	—	—	—	—	—	—	—	—
P test														
H_1: Linear	—	0.151	0.147	0.377	0.072*	0.778	0.029**	0.379	0.089*	0.818	0.529	0.079*	0.009***	0.693
H_1: Tobit	0.492	—	0.038**	0.142	0.009***	0.132	0.050**	0.656	0.017**	0.747	0.216	0.005***	0.002***	0.129
H_1: FRM-Logit	0.297	0.112	—	0.157	0.014**	0.280	0.010***	0.135	0.021**	0.398	0.087*	0.187	0.006***	0.225
H_1: FRM-Probit	0.264	0.055*	0.069*	—	0.014**	0.198	0.010***	0.187	0.023**	0.242	0.094*	0.667	0.003***	0.169
H_1: FRM-Loglog	0.306	0.110	0.058*	0.138	—	0.426	0.013**	0.072*	0.380	0.318	0.066*	0.023**	0.061*	0.319
H_1: FRM-Cloglog	0.064*	0.075*	0.029**	0.048**	0.009***	—	0.010***	0.374	0.012**	0.771	0.493	0.008***	0.001***	0.077*
H_1: GFRM-I-Logit	0.059*	0.044**	0.028**	0.045**	0.009***	0.201	—	0.492	0.011**	0.851	0.102	0.013**	0.000***	0.174
H_1: GFRM-I-Probit	0.777	0.431	0.373	0.430	0.657	0.652	0.143	—	0.736	0.944	0.031**	0.013**	0.001***	0.507
H_1: GFRM-I-Loglog	0.306	0.110	0.058*	0.138	0.434	0.426	0.013**	0.072*	—	0.318	0.066*	0.023**	0.061*	0.319
H_1: GFRM-I-Cloglog	0.009***	0.062*	0.025**	0.041**	0.007***	0.130	0.007***	0.454	0.010***	—	0.156	0.006***	0.000***	0.121
H_1: GFRM-II-Logit	0.320	0.141	0.055*	0.140	0.015**	0.221	0.400	0.624	0.019**	0.450	—	0.023**	0.000***	0.179
H_1: GFRM-II-Probit	0.319	0.455	0.047**	0.417	0.011**	0.204	0.017**	0.273	0.016**	0.563	0.087*	—	0.002***	0.157
H_1: GFRM-II-Loglog	0.102	0.563	0.253	0.481	0.120	0.958	0.107	0.042**	0.246	0.755	0.335	0.009***	—	0.831
H_1: GFRM-II-Cloglog	0.064*	0.075*	0.029**	0.048**	0.009***	0.404	0.010***	0.374	0.012**	0.771	0.493	0.008***	0.001***	—

Note: ***, ** and * denote test statistics which are significant at 1%, 5% or 10%, respectively; FRM = Fractional regression model; GFRM-I = Generalized FRM - type I; GFRM-II = Generalized FRM - type II.

Table 2: Specification tests for the first component of two-part models (p-values)

	FRM				GFRM-I				GFRM-II			
	Logit	Probit	Loglog	Cloglog	Logit	Probit	Loglog	Cloglog	Logit	Probit	Loglog	Cloglog
RESET test	0.145	0.326	0.397	0.105	0.034**	0.046**	0.408	0.036**	0.000***	0.186	0.054**	0.024**
GOFF-I test	0.102	0.410	—	0.251	—	—	—	—	—	—	—	—
GOFF-II test	0.299	0.256	0.365	—	—	—	—	—	—	—	—	—
P test												
H_1: FRM-Logit	—	0.067*	0.014**	0.849	0.003***	0.032**	0.020**	0.039**	0.012**	0.137	0.003***	0.629
H_1: FRM-Probit	0.931	—	0.125	0.964	0.004***	0.037**	0.161	0.039**	0.023**	0.150	0.033**	0.626
H_1: FRM-Loglog	0.432	0.830	—	0.657	0.003***	0.013**	0.791	0.039**	0.002***	0.716	0.000***	0.683
H_1: FRM-Cloglog	0.182	0.024**	0.011**	—	0.005***	0.016**	0.015**	0.039**	0.002***	0.018**	0.002***	0.709
H_1: GFRM-I-Logit	0.057*	0.013**	0.005***	0.132	—	0.031**	0.007***	0.039**	0.000***	0.009***	0.001***	0.053**
H_1: GFRM-I-Probit	0.812	0.398	0.113	0.988	0.004***	—	0.145	0.039**	0.019**	0.463	0.030**	0.547
H_1: GFRM-I-Loglog	0.432	0.830	0.726	0.657	0.003***	0.013**	—	0.039**	0.002***	0.716	0.000***	0.683
H_1: GFRM-I-Cloglog	0.070*	0.019**	0.006***	0.227	0.009***	0.036**	0.008***	—	0.000***	0.013**	0.001***	0.095*
H_1: GFRM-II-Logit	0.344	0.998	0.002***	0.884	0.013**	0.044**	0.003**	0.039**	—	0.749	0.000***	0.692
H_1: GFRM-II-Probit	0.634	0.154	0.082*	0.872	0.004***	0.037**	0.114	0.039**	0.025**	—	0.026**	0.532
H_1: GFRM-II-Loglog	0.278	0.444	0.012**	0.842	0.010***	0.046**	0.021**	0.039**	0.017**	0.366	—	0.649
H_1: GFRM-II-Cloglog	0.182	0.024**	0.011**	0.139	0.005***	0.016**	0.015**	0.039**	0.002***	0.018**	0.002***	—

Notes: ***, ** and * denote test statistics which are significant at 1%, 5% or 10%, respectively; FRM = Fractional regression model; GFRM-I = Generalized FRM - type I; GFRM-II = Generalized FRM - type II.

Table 3: Specification tests for the second component of two-part models (p-values)

	Linear	FRM				GFRM-I				GFRM-II			
		Logit	Probit	Loglog	Cloglog	Logit	Probit	Loglog	Cloglog	Logit	Probit	Loglog	Cloglog
RESET test	0.593	0.736	0.705	0.924	0.516	0.960	0.516	0.821	0.392	0.913	0.718	0.930	0.523
GOFF-I test	—	0.819	0.676	—	0.502	—	—	—	—	—	—	—	—
GOFF-II test	—	0.708	0.746	0.984	—	—	—	—	—	—	—	—	—
P test													
H_1: FRM-Linear	—	0.423	0.402	0.698	0.337	0.414	0.593	0.666	0.335	0.699	0.386	0.640	0.326
H_1: FRM-Logit	0.358	—	0.421	0.886	0.439	0.292	0.273	0.955	0.441	0.896	0.468	0.882	0.444
H_1: FRM-Probit	0.372	0.460	—	0.988	0.485	0.381	0.358	0.920	0.499	0.989	0.747	0.956	0.491
H_1: FRM-Loglog	0.437	0.818	0.663	—	0.444	0.983	0.326	0.258	0.432	0.263	0.671	0.952	0.450
H_1: FRM-Cloglog	0.347	0.710	0.729	0.997	—	0.692	0.901	0.946	0.876	0.936	0.731	0.973	0.193
H_1: GFRM-I-Logit	0.703	0.447	0.990	0.988	0.669	—	0.876	0.907	0.730	0.721	0.871	0.928	0.661
H_1: GFRM-I-Probit	0.284	0.611	0.758	0.892	0.585	0.446	—	0.955	0.575	0.788	0.842	0.897	0.584
H_1: GFRM-I-Loglog	0.437	0.818	0.663	0.443	0.444	0.983	0.326	—	0.432	0.263	0.671	0.952	0.450
H_1: GFRM-I-Cloglog	0.374	0.334	0.363	0.604	0.112	0.183	0.983	0.630	—	0.446	0.320	0.509	0.112
H_1: GFRM-II-Logit	0.404	0.709	0.586	0.205	0.420	0.890	0.295	0.189	0.407	—	0.598	0.521	0.426
H_1: GFRM-II-Probit	0.365	0.504	0.744	0.992	0.486	0.424	0.374	0.924	0.502	0.981	—	0.960	0.492
H_1: GFRM-II-Loglog	0.430	0.865	0.669	0.986	0.453	0.863	0.324	0.906	0.444	0.660	0.677	—	0.459
H_1: GFRM-II-Cloglog	0.347	0.710	0.729	0.997	0.890	0.692	0.901	0.946	0.876	0.936	0.731	0.973	—

Notes: ***, ** and * denote test statistics which are significant at 1%, 5% or 10%, respectively; FRM = Fractional regression model; GFRM-I = Generalized FRM - type I; GFRM-II = Generalized FRM - type II.

Table 4: Estimation results for linear, tobit, logit and selected fractional regression models

	One-part models				Two-part models						
					1st part		2nd part				
	Linear	Tobit	Logit	Cloglog	Logit	Cloglog	Linear	Logit	Probit	Loglog	Cloglog
LANDLORD	0.032	0.080	0.242	0.110	0.540	0.350	-0.028	-0.115	-0.073	-0.076	-0.086
	(0.033)	(0.056)	(0.205)	(0.104)	(0.299)	(0.216)	(0.032)	(0.139)	(0.086)	(0.108)	(0.090)
LIVESTOCK	0.054	0.128^{*}	0.317	0.157	1.149^{**}	0.856^{**}	-0.088^{**}	-0.365^{**}	-0.224^{**}	-0.284^{**}	-0.232^{**}
	(0.041)	(0.070)	(0.218)	(0.121)	(0.422)	(0.345)	(0.036)	(0.152)	(0.094)	(0.117)	(0.100)
CROP	0.071^{*}	0.130^{*}	0.457^{**}	0.201	0.823^{*}	0.626^{*}	-0.002	-0.001	0.000	-0.003	0.002
	(0.041)	(0.069)	(0.228)	(0.124)	(0.420)	(0.347)	(0.038)	(0.166)	(0.102)	(0.132)	(0.104)
SIZE	0.405^{***}	1.130^{***}	4.737^{***}	2.213^{***}	4.255^{**}	2.701^{**}	0.654^{***}	2.898^{***}	1.782^{***}	2.310^{***}	1.822^{***}
	(0.096)	(0.357)	(1.217)	(0.512)	(1.776)	(1.077)	(0.204)	(0.969)	(0.585)	(0.800)	(0.572)
SUBSIDIES	-0.094	-0.163^{**}	-0.451	-0.400	-0.831^{*}	-0.792^{*}	-0.123^{**}	-0.509^{**}	-0.314^{**}	-0.370^{*}	-0.341^{**}
	(0.083)	(0.080)	(0.451)	(0.271)	(0.534)	(0.417)	(0.060)	(0.257)	(0.158)	(0.192)	(0.171)
ALTO	-0.043	-0.040	-0.278	-0.126	0.162	0.104	-0.096^{**}	-0.416^{**}	-0.255^{**}	-0.338^{**}	-0.258^{**}
	(0.037)	(0.071)	(0.241)	(0.120)	(0.382)	(0.281)	(0.039)	(0.172)	(0.106)	(0.138)	(0.107)
CENTRAL	-0.081^{*}	-0.101	-0.527^{*}	-0.251^{*}	-0.102^{*}	-0.058^{*}	-0.136^{***}	-0.589^{***}	-0.361^{***}	-0.464^{***}	-0.371^{***}
	(0.045)	(0.081)	(0.285)	(0.146)	(0.435)	(0.319)	(0.045)	(0.200)	(0.123)	(0.159)	(0.126)
BAIXO	-0.057	-0.091	-0.419	-0.188	-0.211	-0.184	-0.078^{*}	-0.349^{*}	-0.211^{*}	-0.288^{*}	-0.206^{*}
	(0.040)	(0.072)	(0.259)	(0.132)	(0.398)	(0.300)	(0.040)	(0.182)	(0.111)	(0.147)	(0.112)
σ / α	—	0.360^{***}	—	—	—	—					
		(0.023)									
CONSTANT	0.775^{***}	0.820^{***}	1.117^{***}	0.415^{***}	-1.150^{***}	-1.129^{***}	0.744^{***}	1.021^{***}	0.630^{***}	1.152^{***}	0.300^{**}
	(0.060)	(0.088)	(0.339)	(0.189)	(0.532)	(0.423)	(0.053)	(0.236)	(0.144)	(0.186)	(0.149)
Number of observations	266	266	266	266	266	266	149	149	149	149	149
R^2	0.085	0.080	0.086	0.091	0.101	0.105	0.183	0.185	0.185	0.185	0.184
% of predictions outside the unit interval	1.1	—	—	—	—	—	1.3	—	—	—	—

Notes: below the coefficients we report standard errors in parentheses; ***, ** and * denote coefficients which are significant at 1%, 5% or 10%, respectively.

Table 5: Sample averages of partial effects

	One-part models				Two-part models									
					Logit (1st part) +					Cloglog (1st part) +				
	Linear	Tobit	Logit	Cloglog	Linear	Logit	Probit	Loglog	Cloglog	Linear	Logit	Probit	Loglog	Cloglog
LANDLORD	0.032	0.047	0.040	0.035	0.062	0.062	0.062	0.064	0.061	0.052	0.053	0.053	0.055	0.051
LIVESTOCK	0.054	0.076	0.052	0.050	0.119	0.121	0.121	0.121	0.121	0.119	0.121	0.121	0.122	0.122
CROP	0.071	0.077	0.076	0.064	0.112	0.112	0.112	0.112	0.113	0.115	0.115	0.115	0.115	0.116
SIZE	0.405	0.671	0.785	0.702	0.869	0.869	0.868	0.873	0.863	0.785	0.785	0.784	0.789	0.780
SUBSIDIES	-0.094	-0.097	-0.075	-0.127	-0.168	-0.164	-0.164	-0.160	-0.166	-0.200	-0.196	-0.196	-0.192	-0.199
ALTO	-0.043	-0.024	-0.046	-0.040	-0.020	-0.019	-0.019	-0.020	-0.018	-0.023	-0.022	-0.022	-0.024	-0.021
CENTRAL	-0.081	-0.060	-0.087	-0.080	-0.074	-0.072	-0.072	-0.073	-0.071	-0.071	-0.069	-0.069	-0.069	-0.068
BAIXO	-0.057	-0.054	-0.069	-0.060	-0.063	-0.063	-0.063	-0.065	-0.061	-0.068	-0.068	-0.068	-0.070	-0.066

Figure 1: Standard fractional regression models

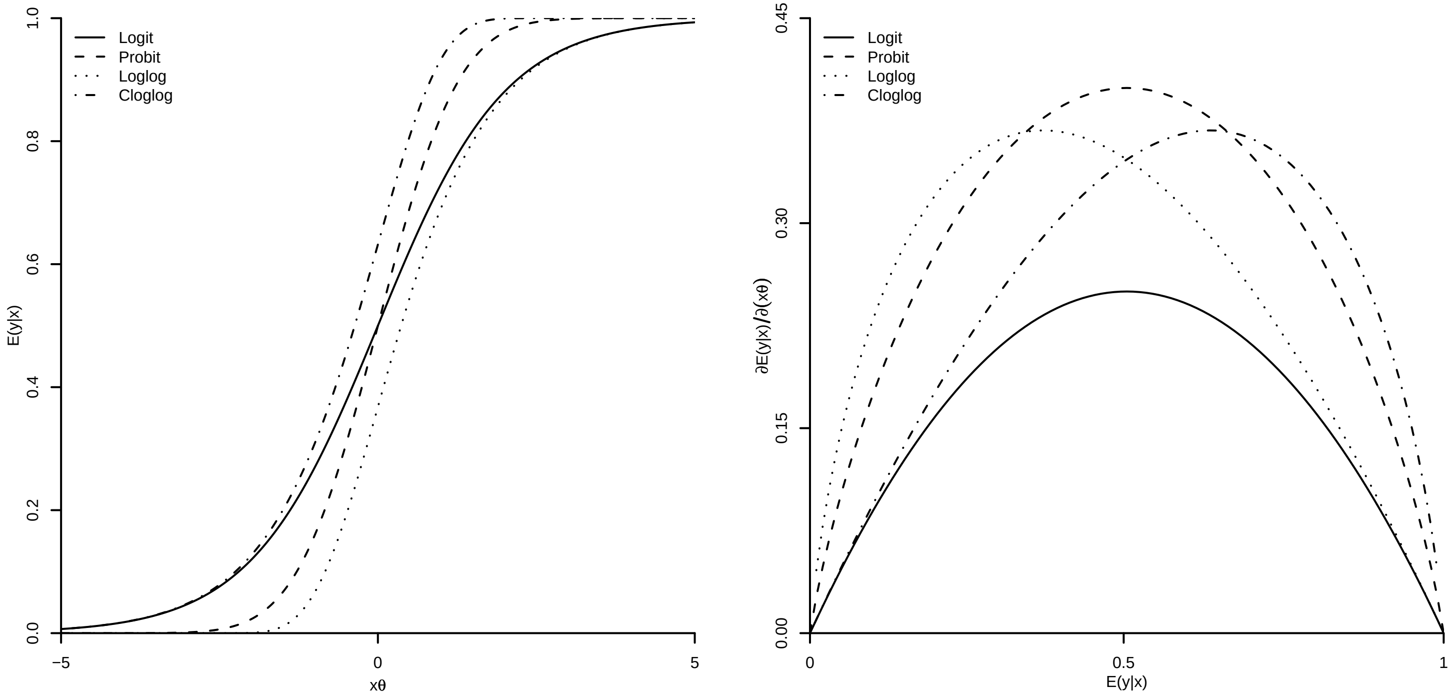

Figure 2: Generalized logit fractional regression models

Type I model

Type II model

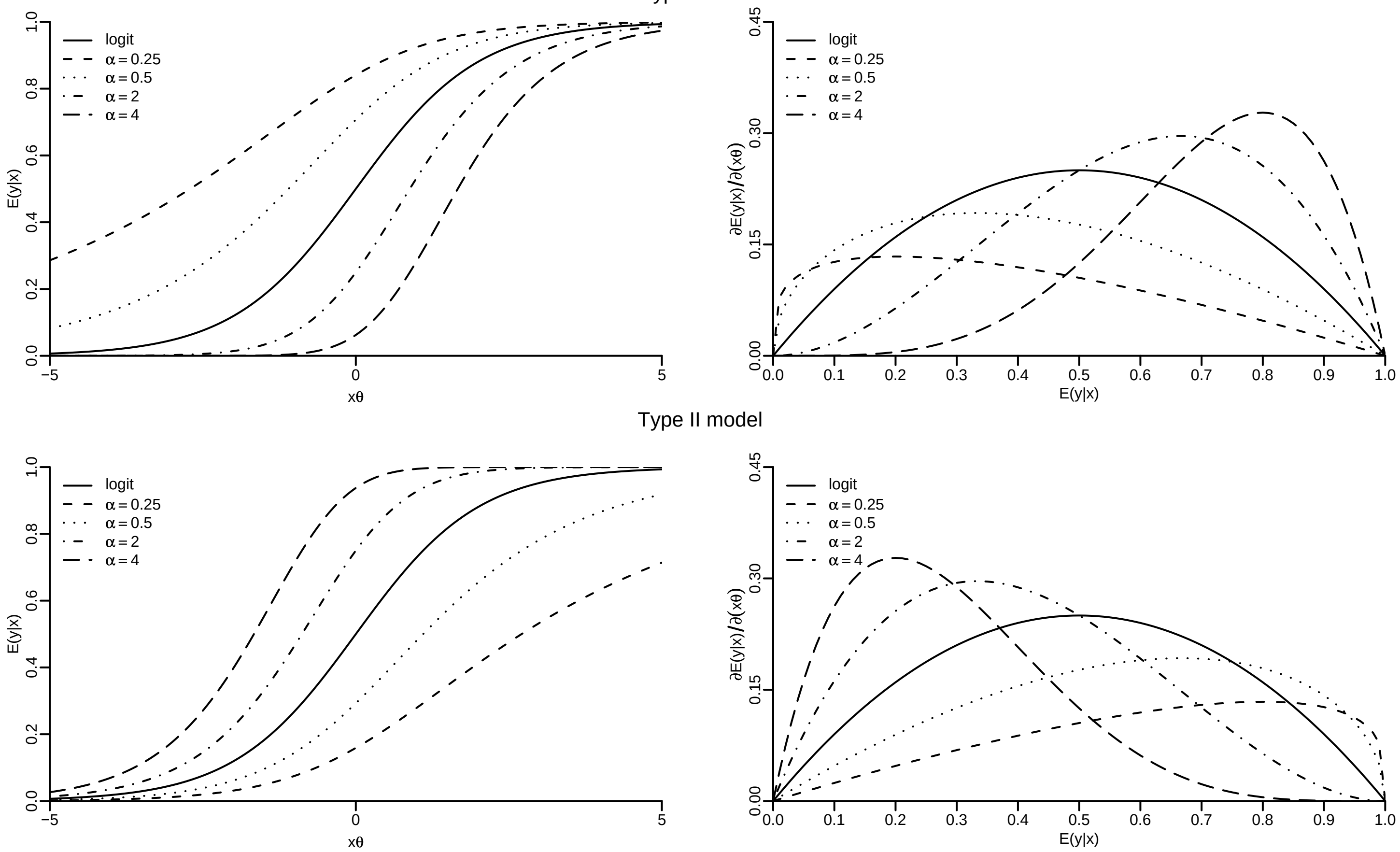

Figure 3: Partial effects for two−part models

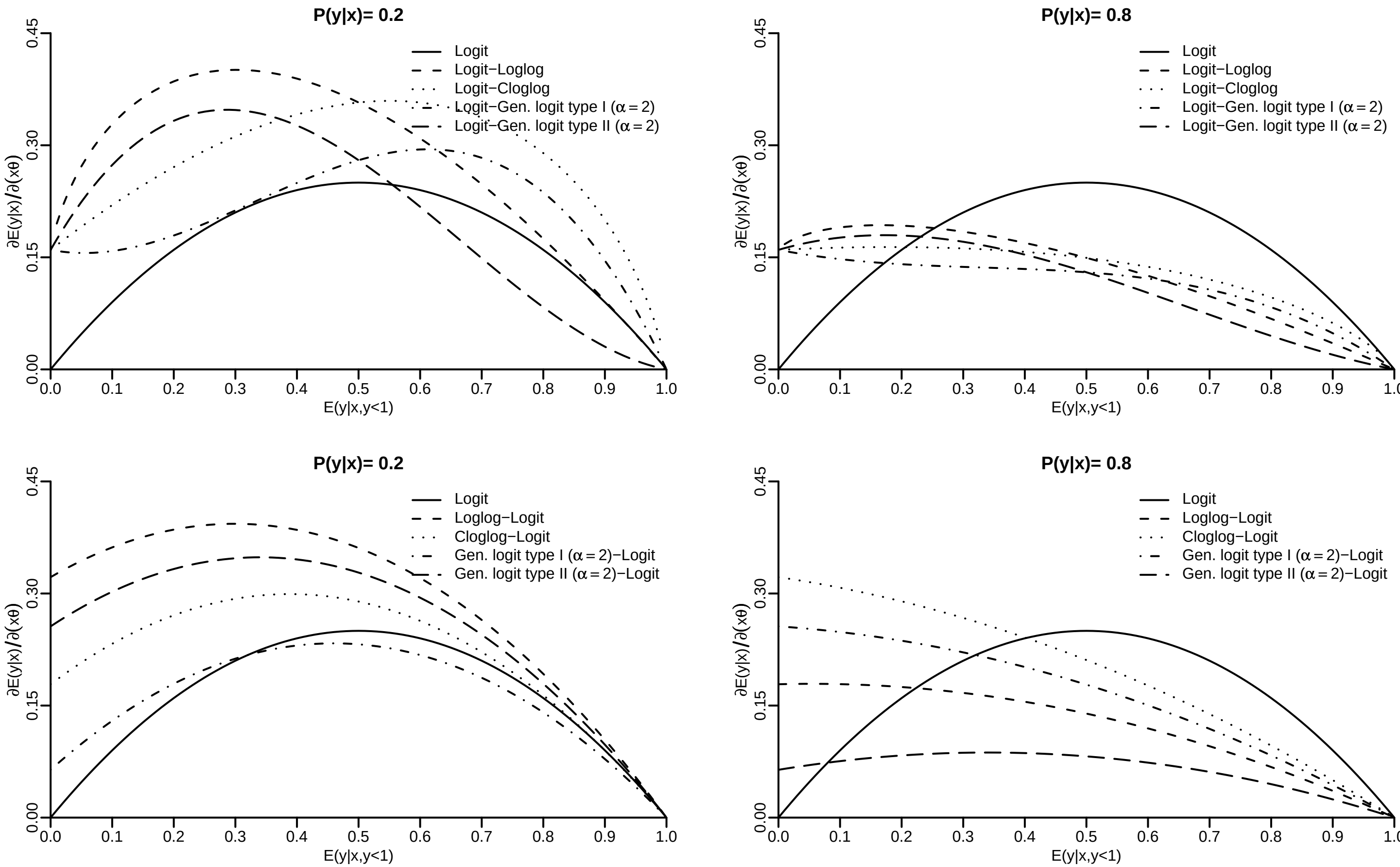

Figure 4: Predicted partial effects and efficiency scores as a function of the SIZE variable