Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

Moment-based estimation of nonlinear regression models with boundary outcomes and endogeneity, with applications to nonnegative and fractional responses

Esmeralda A. Ramalho & Joaquim J. S. Ramalho

To cite this article: Esmeralda A. Ramalho & Joaquim J. S. Ramalho (2017) Moment-based estimation of nonlinear regression models with boundary outcomes and endogeneity, with applications to nonnegative and fractional responses, Econometric Reviews, 36:4, 397-420, DOI: 10.1080/07474938.2014.976531

To link to this article: https://doi.org/10.1080/07474938.2014.976531

Published online: 18 Nov 2015.

ECONOMETRIC REVIEWS
2017, VOL. 36, NO. 4, 397–420
http://dx.doi.org/10.1080/07474938.2014.976531

Taylor & Francis
Taylor & Francis Group

Moment-based estimation of nonlinear regression models with boundary outcomes and endogeneity, with applications to nonnegative and fractional responses

Esmeralda A. Ramalho and Joaquim J. S. Ramalho

Department of Economics and CEFAGE-UE, Universidade de Evora, Evora, Portugal

ABSTRACT
In this article, we suggest simple moment-based estimators to deal with unobserved heterogeneity in a special class of nonlinear regression models that includes as main particular cases exponential models for nonnegative responses and logit and complementary loglog models for fractional responses. The proposed estimators: (i) treat observed and omitted covariates in a similar manner; (ii) can deal with boundary outcomes; (iii) accommodate endogenous explanatory variables without requiring knowledge on the reduced form model, although such information may be easily incorporated in the estimation process; (iv) do not require distributional assumptions on the unobservables, a conditional mean assumption being enough for consistent estimation of the structural parameters; and (v) under the additional assumption that the dependence between observables and unobservables is restricted to the conditional mean, produce consistent estimators of partial effects conditional only on observables.

KEYWORDS
Boundary outcomes; endogeneity; exponential regression; fractional regression; transformation regression models; unobserved heterogeneity

JEL CLASSIFICATION
C25; C51; C26

1. Introduction

Economic theory often postulates that a response variable depends on both observed and unobserved individual factors; see inter alia Heckman (2000, 2001), and his notion of a Marshallian structural function, and Wooldridge (2005), and his related concept of a structural expectation of interest. Therefore, empirical researchers often have to deal with the problem of "omitted variables" or "unobservables" in their econometric models. When the model is linear in the parameters, this issue is easily dealt with. For example, if the omitted variables are uncorrelated with the variables included in the model, then unobservables may be simply ignored, and standard application of ordinary least squares (OLS) produces unbiased estimators of the parameters of interest; if, instead, unobserved and observed covariates are correlated, then, provided that a set of instruments is available for the endogenous regressors, instrumental variables based approaches, such as the generalized method of moments (GMM), may be applied.

While linear models are widely used in econometrics, there are many circumstances where it is preferable to specify a nonlinear regression model, such as when the dependent variable, y, has a bounded nature. In such a case, linear models provide, in general, an inadequate description of the behavior of y, because they do not impose any restriction on the range of values yielded by the structural function relating y to the observables x and the unobservables u. Conversely, in a structural model of the type $y = G(x\theta + u)$, where $G(\cdot)$ is a nonlinear function, it is straightforward to take into account the bounded nature of y. For example, if all realizations of y are nonnegative or fractional, then several alternative nonlinear models that ensure that, respectively, $G(\cdot) > 0$ or $0 < G(\cdot) < 1$, are available. Unfortunately, in the framework of nonlinear models it is much more complicated to deal with unobserved heterogeneity

CONTACT Esmeralda A. Ramalho ela@uevora.pt Department of Economics, Universidade de Évora, Largo dos Colegiais, ÉVORA 7000-803, Portugal.

and its consequences are in general more serious. For example, Ramalho and Ramalho (2010) found the following consequences of neglected heterogeneity in the context of binary and fractional regression models: it produces an attenuation bias in the estimation of regression coefficients; apart from some particular cases, it generates biased estimation of (averaged across the distribution of unobservables) partial effects; and, although innocuous for the size of Wald tests for the significance of observed regressors, it substantially reduces their power.[1]

Despite the general acknowledgement of the deleterious effects of unobserved heterogeneity in nonlinear specifications, many empirical researchers still specify nonlinear models that simply do not allow for unobservables as if ignoring the problem would eliminate it. Other common practice in empirical work (which, in practical terms, is identical to the previous approach) is the introduction of heterogeneity in the model in such a way that it can immediately be discarded again, i.e., observables and unobservables are treated, often without any plausible reason, in a non-symmetrical way, just to ensure the separability of the observable and unobservable components. Authors that do incorporate the heterogeneity in the model in a sensible manner then typically choose one of the following strategies: (i) make strong distributional assumptions for the unobservables, which often generate poorly fitting models; (ii) work with linearized versions of the model of interest (e.g., log-transformed models for nonnegative responses), which, typically, cannot be directly applied in cases where boundary values of y are observed with nonzero probability; or (iii) use nonparametric techniques, which avoid the specification of a functional form for the structural model, but, given their larger complexity, are less appealing to applied researchers than parametric techniques.

In this article we propose a new class of transformation regression models to deal with boundary outcomes, neglected heterogeneity, and endogeneity issues in nonlinear models that treat observed and omitted covariates in a similar manner. Our approach is, on the one hand, less flexible than the three mentioned strategies, since it applies only to a specific class of nonlinear regression models, the most prominent examples of that class being models for nonnegative and fractional responses. However, for these particular models, the proposed approach displays several advantages. First, unlike the linearized models used by strategy (ii), our transformation regression model accommodates values of y observed at one of its boundaries (e.g., the value zero of nonnegative outcomes; the value zero or one of fractional response variables). Second, because the suggested model may be estimated by GMM, its implementation is typically much simpler than those of strategy (i), where often the parameters have to be estimated using simulation techniques, and strategy (iii), where substantial technical and programming skills are often required. Third, unlike strategy (i), no distributional assumptions are required, a conditional mean assumption regarding the unobservables being enough for consistent estimation of the parameters of interest. Fourth, our approach can deal with endogenous covariates without requiring knowledge on the reduced form model, although such information may be easily incorporated in the estimation process. In contrast, most of the approaches following strategies (i) and (iii) above require either the estimation of the reduced form model or heavier assumptions on the relationship between observables and unobservables.

This article focuses on the estimation of the parameters that appear in the structural model, which are of interest in its own right for policy analysis or for testing restrictions imposed by economic theory, for example. However, some authors, notably Wooldridge (2005), argue that in the presence of unobservables, the quantities of primary interest for empirical analysis are often the partial effects averaged across the population distribution of any unobserved heterogeneity. Therefore, in this paper we consider also the estimation of partial effects conditional only on observables and show how, after obtaining consistent estimates for the structural parameters using the proposed transformation regression model, it is also possible to estimate consistently those quantities under the additional assumption that the dependence between observables and unobservables is restricted to the conditional mean.

[1]We use the term "neglected heterogeneity" to designate the case where the unobserved and the included covariates are independent.

All results derived in the paper are first presented in general terms and then specialized for exponential (Wooldridge, 1992) and fractional (Papke and Wooldridge, 1996) regression models. In the former case, several estimators already known in the econometric literature are produced. In the latter case, new estimators for dealing simultaneously with boundary outcomes, neglected heterogeneity and endogeneity issues are obtained. Given the large variety of economic models that have a dependent variable with a fractional nature (e.g., pension plan participation rates, firm market share, proportion of debt in the financing mix of firms, fraction of land area allocated to agriculture, and proportion of exports in total sales), the large number of cases with boundary outcomes (e.g., many firms do not use debt and do not export), and the fact that unobservables virtually affect all econometric models, the proposed transformation regression models are potentially useful for many areas of applied economics.

This article is organized as follows. Section 2 discusses briefly the specification and estimation of nonlinear structural models. Section 3 describes the proposed transformation regression model. Section 4 considers estimation of partial effects conditional on observables. Section 5 uses Monte Carlo methods to compare the finite sample performance of some alternative estimators. Section 6 presents an empirical application concerning the proportion of debt in firms' capital structure. Finally, Section 7 concludes.

2. Specification and estimation of nonlinear regression models

Let y be an observed (limited) dependent variable, and let x and q be k and c vectors of observed explanatory variables and unobserved heterogeneity, respectively. Assume that x contains a constant term and denote by θ and η the vectors of parameters associated to x and q, respectively. Without loss of generality, let $u \equiv q\eta$.

Throughout this article, we consider what Heckman (2000) calls a "well-posed economic model", that is, "a model that specifies all of the input processes, observed and unobserved by the analyst, and their relationship to outputs (p. 47)". We also assume that economic theory implies restrictions on the structure of the model, generating a nonlinear single-index representation for the relationship between y and (x, u). Under these assumptions, the resultant structural model, called by Heckman (2000, 2001) a Marshallian causal function, may be specified as

$$y = G(x\theta + u), \tag{1}$$

where $G(\cdot)$ is a known nonlinear function that imposes the bounded nature of y on the model and is assumed to be strictly monotonic, continuously differentiable, and not additively separable.[2]

From (1), it follows that

$$E(y|x) = E_u[G(x\theta + u)] = \int_{\mathcal{U}} G(x\theta + u) f(u|x)\, du, \tag{2}$$

where $E_u[\cdot]$ denotes expectation with respect to the conditional distribution of u and $\mathcal{U}$ and $f(u|x)$ denote, respectively, the sample space and the conditional (on the observables) density of u, which in this case, for simplicity, is assumed to be a scalar. Equation (2) shows that conditioning on the observed explanatory variables does not remove, in general, the dependency of the model on unobservables (see Section 3.4.1 for a well-known exception). From (2), it follows that the (conditional only on observables) partial effects of unitary changes in a continuous covariate x_l on y are given by

$$\frac{\partial E(y|x)}{\partial x_l} = \int_{\mathcal{U}} \frac{\partial [G(x\theta + u)]}{\partial x_l} f(u|x)\, du, \tag{3}$$

assuming that the integral and differential operators are interchangeable.

The transformation regression model that is proposed in this article, see Section 3, is defined by a set of orthogonality conditions between a function of the unobservables, say $u^* \equiv u^*(y, x; \theta)$, and a set of

[2] See Heckman (2000) for a rigorous definition of Marshallian causal functions.

s instrumental variables, $s \geq k$, which we denote by z:

$$E\left(z'u^*\right) = 0. \tag{4}$$

The instruments z may or may not coincide with the explanatory variables, depending on whether the latter variables may be viewed as exogenous or endogenous. As shown later on, u^* may be a nonlinear function of θ, and therefore, the parameters of interest that appear in (4) have to be estimated by GMM or any other method appropriate for moment condition models. In this article, we consider the so-called two-step GMM estimator defined as

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \hat{m}(\theta)' \hat{\Omega}(\tilde{\theta})^{-1} \hat{m}(\theta), \tag{5}$$

where Θ denotes the parameter space, $\hat{m}(\theta) \equiv \sum_{j=1}^{N} m(y_j, x_j, z_j; \theta) / N$, $m(\cdot) = z'u^*$, j indexes each sampling unit, N denotes the number of observations, $\hat{\Omega}(\theta) \equiv \sum_{j=1}^{N} m_j(\theta) m_j(\theta)' / N$, and $\tilde{\theta}$ is some preliminary estimator defined by an equation similar to (5) but with $\hat{\Omega}(\theta)$ replaced by the identity matrix.

3. Transformation regression models

Estimating θ directly from (1) or (2) is typically a challenging task given the presence of unobservables in both expressions. One possible alternative is the use of transformation regression models. In this section, we first briefly review the typical transformation that has been applied to most nonlinear models. Then, we propose a new transformation regression model that circumvents the inability of the standard approach to accommodate boundary values of the response variable. Both approaches transform the original model (1) in such a way that the coherence with the economic theory that implied Eq. (1) is kept, but orthogonality conditions of the type given by (4) may be straightforwardly generated under the assumption that

$$E(u^* | z) = 0. \tag{6}$$

3.1. The standard linearization approach

Assume that there is a monotonic function $H(\cdot) = G(\cdot)^{-1}$ that, applied to both sides of (1), gives rise to the linear model

$$H(y) = x\theta + u. \tag{7}$$

If u was observed, Eqs. (1) and (7) would represent exactly the same deterministic relationship and it would be indifferent to work with either equation. However, given that u is not observed and is additively separable only in (7), identification and estimation of θ becomes much simpler than in model (1). Indeed, assuming that $E(u|z)$ is a constant not depending on z, consistent GMM estimators for the structural parameters are straightforwardly obtained by considering orthogonality conditions generated from (6), with $u^* = u$

$$E\left\{\left[H(y) - x\theta\right] | z\right\} = 0. \tag{8}$$

When $z = x$, this corresponds to a simple estimation of (7) by OLS.

The transformation regression model defined by (7) is very simple and may be applied to a wide variety of nonlinear regression models. However, the $H(\cdot)$ function is often not defined for boundary values of y, as the examples of Section 3.4 illustrate.

3.2. The proposed transformation

To overcome the problem associated to the linearized model (7), we propose next a different transformation of the structural model (1). This new transformation is slightly more complicated than the previous one and gives rise to a regression model that, similarly to the original model, is nonlinear in the parameters. However, the problems with the boundary values that affect the linearized regression model are in general attenuated and, in some cases, even eliminated.

Assume that the function $G(\cdot)$ in (1) may be decomposed as

$$G(x\theta + u) = G_1\left[G_2(x\theta + u)\right], \tag{9}$$

where $G_1(\cdot)$ is an invertible function and $G_2(\cdot)$ is a nonlinear function multiplicatively separable into $k + c$ terms, which, for our purposes, is convenient to group into two terms, one function of $x\theta$, and the other function of u, as follows:

$$G_2(x\theta + u) = G_2(x\theta)\, G_2(u). \tag{10}$$

Typically, $G_2(\cdot)$ will be the exponential function. Assume that $G_2(x\theta) \neq 0$ and that $E\left[G_2(u)|z\right]$ is a constant not depending on z. In particular, given that x contains a constant term, we may assume without any loss of generality that $E\left[G_2(u)|z\right] = 1$.[3]

Let $H_1(\cdot) = G_1(\cdot)^{-1}$. Then, from (8), it follows that

$$H_1(y) = G_2(x\theta)\, G_2(u). \tag{11}$$

Dividing both sides of (11) by $G_2(x\theta)$, so that (functions of) x and u become additively separable, and subtracting 1 to both sides of the resultant model produces

$$\frac{H_1(y)}{G_2(x\theta)} - 1 = G_2(u) - 1. \tag{12}$$

As, under the assumptions stated above, $E\left\{\left[G_2(u) - 1\right]|z\right\} = 0$, the left hand-side of this equation may be interpreted as the residual function that appears in (6). Hence, assuming standard rank conditions for identification, consistent GMM estimators for θ may be obtained based on a set of orthogonality functions generated from

$$E\left\{\left[\frac{H_1(y)}{G_2(x\theta)} - 1\right]\middle| z\right\} = 0. \tag{13}$$

The transformation regression model (12) applies to a more restrict class of models than the simple linearized model (7), since it involves more requirements on the definition of $G(\cdot)$. However, note that while $H_1(\cdot)$ merely transforms $G(\cdot)$ into a (possibly nonlinear) function $G_2(\cdot)$ of $x\theta + u$, $H(\cdot)$ goes one step further and reduces $G(\cdot)$ to $x\theta + u$. Thus, as the examples in Section 3.4 illustrate, $H_1(\cdot)$ creates less restrictions in the domain of y, being well defined for (some of) its boundary values.

While the transformation models (7) and (12) represent the same deterministic relationship, their stochastic versions (8) and (13) are not in general equivalent. In particular, the assumptions required for consistent estimation of the parameters of interest in each model do not imply each other. That is, it may be the case that $E(u|z)$ is a constant not depending on z, as required by (8), but that $E\left[G_2(u)|z\right]$ is a function of z, unlike required by (13), and vice-versa. Only under the stronger assumption of statistical independence between z and u will the two models produce simultaneously consistent estimators for the slope parameters. Therefore, a major discrepancy between the estimated parameters might suggest that some form of misspecification is present in one of the models. Although in this paper we do not consider the development of specific criteria for choosing between the two models, in the Monte Carlo section we investigate the ability of the popular RESET test to detect this type of misspecification.

[3]For example, if $G_2(\cdot)$ is an exponential function, then we may redefine the constant θ_0 and the error term as $\beta_0 = \theta_0 + \log E\left[\exp(u)\right]$ and $\varepsilon = \exp(u)/E\left[\exp(u)\right]$, respectively.

3.3. An alternative GMM estimator for the case of endogenous explanatory variables

By defining the composition of the vector z appropriately, the GMM estimator proposed in the previous section is valid under a variety of situations, including cases of endogenous covariates. From now on, we denote by GMM_x the estimator that uses $z = x$ and, thus, is only appropriate when endogeneity is not a problem; and by GMM_z the estimator based on a set of instruments z that were chosen in such a way that the consistency of the parameter estimators is achieved also under endogeneity.

As it is clear from (13), a very attractive feature of the GMM_z estimator is that no assumptions about the reduced form of the endogenous explanatory variables need to be made. Moreover, our estimator applies in exactly the same way irrespective of the endogenous regressors being discrete or continuous. Thus, the GMM_z estimator may be seen as a generalization of Mullahy's (1997) estimator for exponential regression models (see Section 3.4.1) and is in clear contrast to most instrumental variable estimators that have been proposed for nonlinear regression models (e.g., Smith and Blundell, 1986; Rivers and Vuong, 1988; Wooldridge, 1997), which are not robust to misspecification of the reduced form of the endogenous covariates and typically require different procedures according to the characteristics of those variables. Nevertheless, a potentially more efficient estimator may be constructed in the presence of reliable information about the reduced form of the endogenous explanatory variables. Next, we outline how such information may be incorporated in the estimation process of the parameters of the transformation model (12) in order to obtain an estimator that is similar in spirit to that suggested by Smith and Blundell (1986) for censored regression models, Rivers and Vuong (1988) for binary response models and Wooldridge (1997) for count data/exponential regression models.

Assume that there are k_1 and k_2 exogenous (x_1) and endogenous (x_2) explanatory variables, respectively. Assume also that strictly monotonic transformations of each x_{2l}, $l = 1, \ldots, k_2$, can be found so that a linear reduced form with additive disturbances can be found. Let $S(x_2)$ denote the vector of those monotonic transformations. Then, we may write

$$S(x_2) = z\pi + v, \tag{14}$$

where z contains x_1, π is an $s \times k_2$ matrix of reduced form parameters, and v is a k_2 vector of reduced form errors. Finally, assume that (u, v) is independent of z and that

$$u = v\rho + \epsilon, \tag{15}$$

where ϵ is independent of v. Under these assumptions, it follows from (1) and (15) that

$$y = G(x\theta + v\rho + \epsilon), \tag{16}$$

and, hence,

$$\frac{H_1(y)}{G_2(x\theta + v\rho)} - 1 = G_2(\epsilon) - 1. \tag{17}$$

Using standard arguments from two-step estimation, it may be shown that GMM estimation based on

$$E\left\{ \left[\frac{H_1(y)}{G_2(x\theta + v\rho)} - 1 \right] \middle| x, v \right\} = 0, \tag{18}$$

with v replaced by $\hat{v} = S(x_2) - z\hat{\pi}$, where $\hat{\pi}$ is an OLS estimator, produces consistent estimators for θ and ρ; see Newey and McFadden (1994) for the consistency of two-step estimators. Alternatively, we may append the first-order conditions for $\hat{\pi}$ to the moment conditions generated from (18) and estimate simultaneously π, θ, and ρ by GMM, which has the advantage of providing directly correct standard-errors to all parameters.

Clearly, unlike the Mullahy-type GMM_z estimator, the consistency of this alternative estimator, denoted from now on by GMM_{xv}, depends crucially on the correctness of both Eq. (15) and the reduced form (14). However, if both equations are correctly specified, then, by using that extra information, the GMM_{xv} estimator is more efficient. Moreover, testing for endogeneity is simpler in this framework: simply test for the significance of the parameters ρ in (18) using any classical GMM test of parametric restrictions; see inter alia Wooldridge (1997) for a discussion of similar tests of endogeneity and Newey and West (1987) for GMM tests of parametric restrictions. In contrast, in the GMM_z framework, it appears that the only form of testing for endogeneity of x_2 would involve the implementation of a standard Hausman test contrasting GMM_x and GMM_z estimators.

3.4. Examples

To illustrate the main results of the article, all results will now be specialized to exponential and fractional regression models. These are two clear examples of models to which the proposed transformation will be particularly useful, since in both cases it is very common to observe the boundary values of zero (exponential and fractional models) or one (fractional models).

3.4.1. Exponential regression model

The exponential model, which is commonly used to describe nonnegative outcomes, may be expressed as

$$G\left(x\theta + u\right) = \exp\left(x\theta + u\right); \tag{19}$$

see, for example, Wooldridge (1992). Although nonlinear, $G(\cdot)$ is multiplicatively separable in terms of x and u, which implies that $E\left(y|x\right) = \exp\left(x\theta\right)$ under the assumption that $E\left[\exp(u)|x\right] = 1$. In this context, conventional application of quasi-maximum likelihood (QML) methods yields consistent estimators for θ; see Santos Silva and Tenreyro (2006) for alternative QML estimators for exponential regression models.

The linearization of model (19) is based on the log-transformation $H(y) = \log(y)$. Because this transformation is not defined for $y = 0$, this approach can only be directly applied to positive data. This is an important limitation of this approach because the excess of zeros is an endemic problem with nonnegative responses, be they discrete (e.g., count data, see Cameron and Trivedi, 2005, pp. 681–682) or continuous (e.g., gravity equations, see Santos Silva and Tenreyro, 2006). Typically, the extensive literature using log-transformed models has overcome this limitation by adding an arbitrary constant to all observations of y or by dropping observations with $y = 0$. As shown by Santos Silva and Tenreyro (2006), both approaches may originate large biases in the estimation of the parameters of interest.

The transformation regression model proposed in this article applies to (19) by defining $G_1(\cdot)$ and $G_2(\cdot)$ as, respectively, a linear and an exponential function. Hence, $H_1(y) = y$ and $\frac{H_1(y)}{G_2(x\theta)} = \frac{y}{\exp(x\theta)}$. This implies that the transformation regression model defined by (12) includes as particular cases two estimators well known in the econometrics literature: the estimator proposed by Mullahy (1997) to deal with endogeneity in count data models; and the Gamma-based QML estimator considered by Manning and Mullahy (2001) and Santos Silva and Tenreyro (2006) for the case of exogenous variables. Clearly, in contrast to the standard linearization approach, in this context there is no problem in dealing with zero outcomes of y.

3.4.2. Fractional regression models

Models for variables defined on the unit interval ($0 \leq y \leq 1$) were first suggested by Papke and Wooldridge (1996); see also the recent survey by Ramalho et al. (2011). In this context, some popular choices for $G(\cdot)$ are the probit, logit, and complementary loglog functional forms described in Table 1. In contrast to the previous example, the terms involving x and u are not directly separable in any of those models. Therefore, as discussed by Ramalho and Ramalho (2010), even under neglected heterogeneity does the Bernoulli-based QML method usually applied in this framework yield inconsistent estimators for θ.

Table 1. Alternative fractional regression models.

Model	$G(x\theta+u)$	$H(y)$	$G_1[G_2(\cdot)]$	$G_2(a)$	$H_1(y)$	$\frac{H_1(y)}{G_2(x\theta)}$
Probit	$\Phi(x\theta+u)$	$\Phi^{-1}(y)$	–	–	–	–
Logit	$\frac{\exp(x\theta+u)}{1+\exp(x\theta+u)}$	$\ln\frac{y}{1-y}$	$\frac{G_2(\cdot)}{1+G_2(\cdot)}$	$\exp(a)$	$\frac{y}{1-y}$	$\frac{y}{1-y}\exp(-x\theta)$
Complementary loglog	$1-\exp[-\exp(x\theta+u)]$	$\ln\left[-\ln(1-y)\right]$	$1-\exp[-G_2(\cdot)]$	$\exp(a)$	$-\ln(1-y)$	$-\ln(1-y)\exp(-x\theta)$

The linear transformation $H(\cdot)$, also given in Table 1, is not defined for both the boundaries values 0 and 1 for the three mentioned fractional regression models. Therefore, to deal with boundary values in linearized fractional models, the same "solutions" described for linearized exponential regression models are in general used and the same criticisms apply. In contrast, the new transformation (12), which can be applied to logit and complementary loglog models, can accommodate one of the two boundary values of fractional responses, see Table 1. This is particularly relevant because most samples cluster only at zero or one (see, for example, the applications by Ramalho and Silva, 2009, and Ramalho et al., 2010, respectively). Moreover, note that we can always redefine the response variable and decide to model its complementary, which means that the transformed logit and complementary loglog models may be used irrespective of the boundary value that is observed with a nonzero probability.

3.5. Comparison with non- and semiparametric approaches

We assume throughout this article that the $G(\cdot)$ function is known (and, hence, $H(y)$ and $H_1(y)$ are also known). Alternatively, y in (1) could be specified simply as a single index model, with $G(\cdot)$ unspecified. Next, we compare the transformation model proposed in this article and some non- and semiparametric approaches considered in the econometrics literature.

Several authors proposed nonparametric estimators for the model $H(y) = x\theta + u$ given in (7), where $H(\cdot)$ is an unknown function and u has unknown distribution function. In particular, Horowitz (1996); Ye and Duan (1997) and Chen (2002), assuming the availability of a $\sqrt{n}$-consistent nonparametric estimator for θ, developed $\sqrt{n}$-consistent estimators for $H(\cdot)$. However, their methods are unable to provide consistent estimators of $E\left(y|x\right)$ for all y and u and require independence between the covariates and the error term. In contrast, in this article we focus on consistent estimation of θ and can handle endogeneity through the use of instruments.

On the other hand, several nonparametric estimators for θ have been proposed, such as the average derivative estimator of Horowitz and Hardle (1996), the maximum rank correlation estimator of Han (1987), the monotone rank estimator of Cavanagh and Sherman (1998) and the pairwise-difference rank estimator of Abrevaya (2003). Again, none of these methods can deal with endogenous explanatory variables. Moreover, because $H(\cdot)$ is left unspecified, θ is only identified up-to-scale and without location.

Semi- and nonparametric methods that allow for endogenous covariates in model (7) have also been proposed, but typically, in contrast to our proposal, such approaches require the continuity of the endogenous variables (e.g., Vanhems and van Keilegom, 2011) and often also of the instrumental variables (e.g., Fève and Florens, 2010). Other disadvantage is that often the rate of convergence of the estimators is no longer the usual $\sqrt{n}$ rate (Fève and Florens, 2010). As far as we know, the only nonparametric approach that allows for both continuous and discrete endogenous and instrumental variables is that by Abrevaya et al. (2010). However, the proposed method merely identifies the sign of the endogenous regressors.

Finally, irrespective of the advantages and disadvantages of any particular nonparametric estimator relative to our transformation regression model, note that most practitioners still prefer working with parametric transformations, since they are easier to implement and interpret than nonparametric ones. In particular, to the best of our knowledge, no applications of nonparametric estimators for fractional responses have been performed yet. Therefore, particularly for empirical researchers working with fractional data, our approach is potentially very useful, since it will allow them to deal simultaneously with boundary values, neglected heterogeneity and endogeneity in a very straightforward way.

4. Estimation of partial effects conditional on observables

As it is clear from (3), estimation of partial effects conditional on observables will require in general making distributional assumptions on the unobservables. Moreover, the integrals that appear in (3), in general, cannot be calculated analytically and have to be computed by numeric integration or simulation. There are, however, some exceptions to this situation. One concerns the exponential regression model, since in this case the structural function $G(\cdot)$ is multiplicatively separable in terms of x and u, see Section 3.4.1, and, thus, $f(u|x)$ does not need to be specified. The other exceptions concern very special combinations of the structural function and the distribution of unobservables, still requiring making distributional assumptions on unobservables but avoiding the computation of integrals. In fact, when u has a specific distribution (typically, the normal distribution), $G(\cdot)$ has a specific form (e.g., a binary probit model), and u is independent of x, or independent conditionally on a set of additional controls, then estimating the naive model that ignores the presence of unobserved heterogeneity (i.e., estimating the misspecified model $E\left(y|x\right) = G\left(x\theta\right)$ instead of the correct model defined in (2)), or the model that adds a set of controls to the index function but still omits u, produces biased estimates for the structural parameters but consistent estimators for the partial effects conditional on unobservables; see Wooldridge (2005) for details and examples and Ramalho and Ramalho (2010) for further discussion.

Wooldridge's (2005) approach still requires making distributional assumptions on u, has the undesirable feature of yielding inconsistent estimators for the structural parameters, and is not generally applicable. The transformation regression models proposed in the previous section are also not generally applicable but, in addition to produce consistent estimators for the structural parameters, have also the ability of generating consistent estimators for conditional partial effects without requiring the full specification of $f(u|x)$, provided that we restrict the dependence between observables and unobservables to the conditional mean (i.e., $E\left\{\left[G_2(u) - 1\right]|x\right\}$ may depend on x but other functions of u not). Under this additional assumption, the following two-step procedure may be used for estimating partial effects for individual i:

1. Obtain the GMM estimator $\hat{\theta}$ and the residuals $\hat{u}_j^* = \frac{H_1(y_j)}{G_2\left(x_j\hat{\theta}\right)} - 1$ and $\hat{u}_j = G_2^{-1}\left(\hat{u}_j^* + 1\right)$, $j = 1, \ldots, N$, where the function G_2 is assumed to be invertible;
2. Compute the partial effect (3) using its sample analog

$$\left(\widehat{\frac{\partial E\left(y|x\right)}{\partial x_l}}\right) = \frac{1}{N}\sum_{j=1}^{N}\frac{\partial\left\{G_1\left[G_2\left(x_i\hat{\theta}\right)G_2\left(\hat{u}_j\right)\right]\right\}}{\partial x_l} = \frac{1}{N}\hat{\theta}_l\sum_{j=1}^{N}g\left(x_i\hat{\theta} + \hat{u}_j\right), \tag{20}$$

where $g(\cdot)$ is the derivative of $G(\cdot)$.

Note that this two-step procedure is valid for all the GMM estimators defined in the previous section: GMM_x, GMM_z, and GMM_{xv}.

The estimator defined in (20) is a natural extension of the smearing technique suggested by Duan (1983) for the log-transformed model, estimating the unknown error distribution by the empirical distribution function of the GMM residuals calculated in step 1. Although rarely used in the economics literature, this is a very simple method to employ in practice. Of course, the variance of (20) will have to be computed using the delta method or the bootstrap, but that is the standard procedure when working with partial effects in nonlinear models. In the former case, the following formula should be used for computing the variance of the partial effects:

$$Var\left[\left(\widehat{\frac{\partial E\left(y|x\right)}{\partial x_l}}\right)\right] = \frac{\partial\left(\partial \widehat{E\left(y|x\right)}/\partial x_l\right)}{\partial\theta}Var\left(\hat{\theta}\right)\frac{\partial\left(\partial \widehat{E\left(y|x\right)}/\partial x_l\right)}{\partial\theta}'; \tag{21}$$

see Abrevaya (2002). In the latter case, an appropriate bootstrap for GMM estimators must be applied (see, e.g., Ramalho, 2006), with the variance of (20) being given by the sample variance of the B estimates of $\left(\frac{\partial E(y|x)}{\partial x_l}\right)$ obtained in the bootstrap samples.

5. Monte Carlo simulation study

In this section, we carry out a Monte Carlo study to investigate the finite-sample performance of the estimators proposed in the article under different simulated scenarios. For comparative purposes, in addition to the GMM_x, GMM_z, and GMM_{xv} estimators, we include also three estimators based on the linearized model (7), and two *QML* estimators. The linearized estimators, denoted for simplicity by LIN_x, LIN_z and LIN_{xv}, are constructed in a similar way to their *GMM* counterparts. Regarding the *QML* estimators, we considered the standard Bernoulli-based *QML* estimator proposed by Papke and Wooldridge (1996) for fractional regression models (denoted by QML_x), which does not account for any type of unobserved heterogeneity; and a variant of that estimator (denoted by QML_{xv}), which was proposed by Wooldridge (2005) to deal with endogeneity issues and is constructed in a similar way to the GMM_{xv} estimator, but, unlike the latter, does not allow for other sources of heterogeneity.

This Monte Carlo study considers three distinct experimental designs, all of them concerning the estimation of a logit fractional regression model. In the first set of experiments, the data is generated without boundary observations, a setting where all estimators discussed in the article may be applied with no need for *ad-hoc* adaptations. We use these experiments to perform a comprehensive analysis of the ability of each method to estimate structural parameters and conditional partial effects. In the second set of experiments, we generate data with boundary observations that result from rounding errors in order to illustrate the advantages of the estimators proposed in this article over the (modified) linearized estimators that are typically applied in such a context. Finally, in the third set of experiments, we generate the data in such a way that the two transformation regression models discussed in the article cannot yield simultaneously consistent estimators for the slope parameters, the aim being the analysis of the ability of the RESET test to detect the misspecification that affects one of the models.

All experiments were repeated 5,000 times using the statistical package *R*. For the nonlinear estimators, the general-purpose optimization function *nlminb*, which is based on a quasi-Newton algorithm, was used in both the GMM and QML cases. Also in both cases, the true values of the parameters were used to initiate the algorithm.

5.1. Experiments without boundary observations

5.1.1. Design

The experiments without boundary observations are based on the structural model

$$y = G(\theta_0 + \theta_1 x_1 + u), \tag{22}$$

where $G(a) = \exp(a) / \left[1 + \exp(a)\right], \theta_0 = 0, \theta_1 = 1$, and x_1 denotes a single covariate. The explanatory variable x_1 is generated from either

$$x_1 = z\pi_1 + v \tag{23}$$

or

$$\ln \frac{x_1}{1 - x_1} = z\pi_2 + v, \tag{24}$$

where z is an $(s-1)$-vector of instrumental variables, which are generated as $\mathcal{N}(0,1)$ random variables, with the elements of z independent of each other and of u and v. The reduced form parameters π_1 and π_2 equal an $(s-1)$-vector of ones times a scalar constant Π_1 and Π_2, respectively. We set $s = \{3, 12\}$, $\Pi_1 = 0.5$, and $\Pi_2 = 1$.

We generate the error terms (u, v) as correlated: their joint distribution is $\mathcal{N}(\mu, \Sigma)$, where $\mu = (-0.5, 0)$ and $\Sigma \in \mathcal{R}^{2\times 2}$ with diagonal elements equal to unity and off-diagonal elements ρ_{uv}. Setting the mean of u equal to -0.5 ensures that $E\left[\exp(u)|z\right] = 1$, as assumed by our *GMM* estimators. Because $E(u|z) \neq 0$ but does not depend on z, *LIN* estimators may be consistent for θ_1 but not for θ_0.

As the error terms have a joint multivariate normal distribution, we may write

$$u = \rho_{uv} v + \epsilon, \tag{25}$$

with ϵ independent of v and z, as in Eq. (15). Moreover, the variance of ϵ is given by $\sigma_\epsilon^2 = 1 - \rho_{uv}^2$. Hence, while in the $\rho_{uv} = 0$ ($\sigma_\epsilon^2 = 1$) case x_1 is exogenous but there is a large amount of neglected heterogeneity, for $\rho_{uv} = \pm 1$ ($\sigma_\epsilon^2 = 0$) x_1 is strongly endogenous but the impact of neglected heterogeneity is irrelevant (if ignored, only the estimation of the parameter θ_0 is affected). In order to measure the effect of different degrees of endogeneity and neglected heterogeneity over the various estimators, we set $\rho_{uv} = \{-1, -0.8, \ldots, 1\}$. In all cases, Monte Carlo samples of size $N = \{200, 1000\}$ are generated.

5.1.2. Estimation of structural parameters

Figure 1 presents the results obtained for the first set of experiments, where the reduced form (23) was used both for generating the data and for estimating purposes. For each experiment, we report for six alternative estimators of the structural parameter θ_1 the following statistics: the mean across replications; the root mean squared error (RMSE); and the empirical coverage of a 95% confidence interval, which was estimated by taking the proportion of cases where the confidence interval covers the true value of θ_1. In order to estimate correctly the standard errors necessary for the computation of the confidence interval, in the implementation of the QML_{xv} and GMM_{xv} estimators both the parameters of the structural and reduced forms were estimated simultaneously. To facilitate the reading of the figures, the results for LIN_x and LIN_{xv} are not reported, but their performances relative to LIN_z are similar to the performances of GMM_x and GMM_{xv} relative to GMM_z, respectively.

The first column of Fig. 1 shows clearly that, irrespective of the value of ρ_{uv}, GMM_{xv}, GMM_z, and LIN_z provide consistent estimation of θ_1. In contrast, all the other estimators are biased in most cases. The GMM_x estimator is consistent only for $\rho_{uv} = 0$, its bias increasing as the degree of endogeneity (in absolute value) increases. The QML_{xv} estimator is consistent only when the neglected heterogeneity can be ignored ($\rho_{uv} = \pm 1$), its bias increasing as the variance of v increases (ρ_{uv} decreases), achieving a maximum for $\sigma_\epsilon^2 = 1$. That is, the QML_{xv} estimator displays the classic attenuation bias that is often mentioned as the main consequence of neglected heterogeneity. As in all simulated cases there are endogeneity and/or neglected heterogeneity, the standard QML_x estimator, used in most empirical applications, displays large biases most of time, except in a particular situation where the effects of endogeneity and neglected heterogeneity seem to compensate each other. Because they do not account for endogeneity, the bias of both the QML_x and GMM_x estimators may be positive or negative, depending on the value of ρ_{uv}.

The analysis of the RMSE of each estimator shows the importance of using additional information in the estimation process in order to obtain more efficient estimators, especially for smaller sample sizes and when less moment conditions are used.[4] Clearly, in the presence of unobserved heterogeneity, if the empirical researcher knows for sure that endogeneity is not an issue, then he/she should use the GMM_x (or LIN_x) estimator; if neglected heterogeneity is not a problem and the reduced form of the endogenous explanatory variable is known, then the QML_{xv} estimator is probably the best option; if the data are affected by both neglected heterogeneity and endogeneity issues and the reduced form of the endogenous regressor is known, then it is preferable to apply the GMM_{xv} (or LIN_{xv}) estimator. On the other hand, the *LIN* estimators display clearly a better RMSE performance than their *GMM* competitors for $N = 200$ and $s = 3$, but that advantage becomes much less important as the sample size and the number of moment conditions increase.

[4]Note that as π_1 is fixed independently of the number of instruments, more instruments imply a higher overall fit of the instruments to the endogenous regressor x_1.

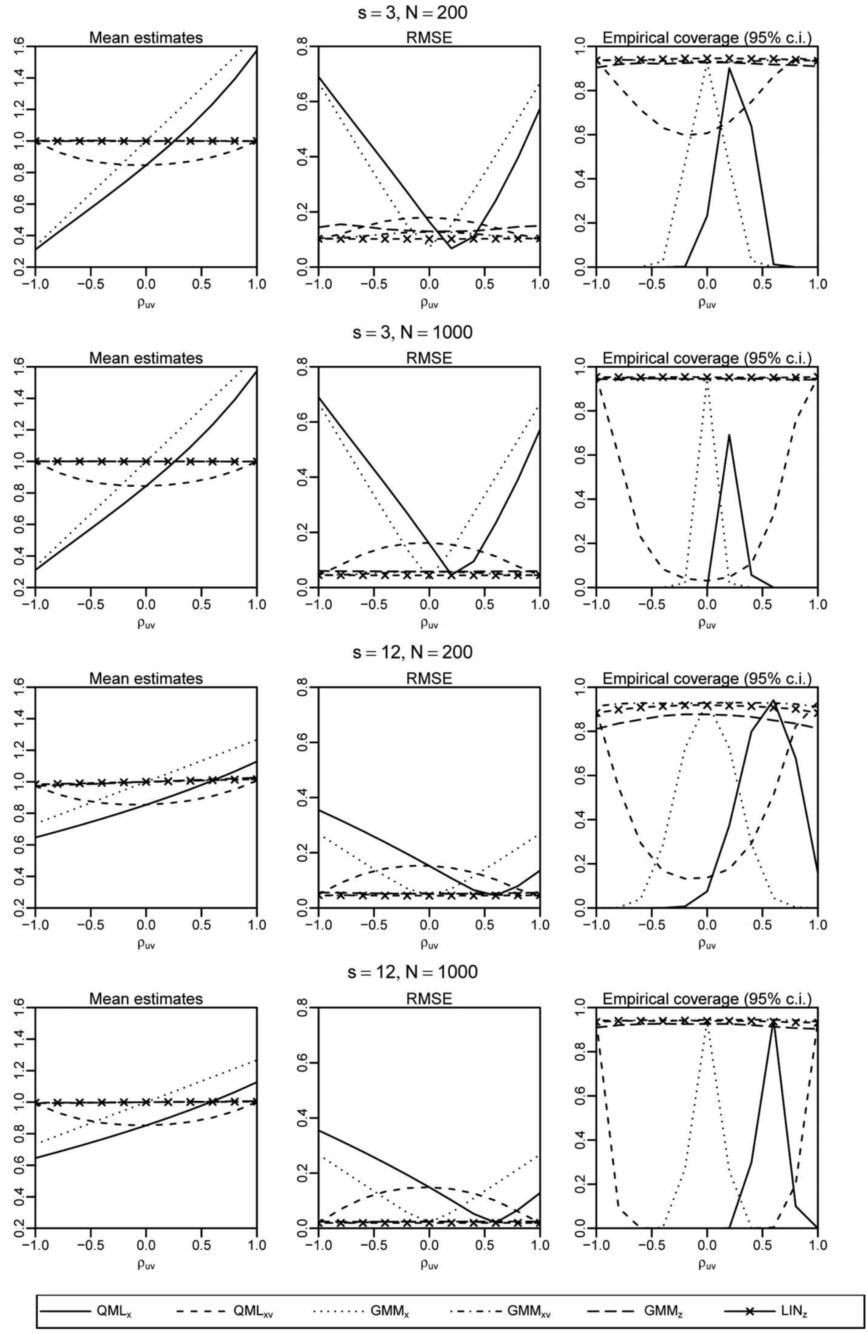

Figure 1. Monte Carlo statistics for the structural parameter θ_1.

Similar conclusions are achieved if we analyze the graphs relative to the coverage of confidence intervals. The last column of Fig. 1 also illustrates the danger of not accounting for the correct type of unobserved heterogeneity: the empirical coverage of the confidence intervals yielded by QML_x, QML_{xv}, and GMM_x tend to zero except in the particular cases of neglected heterogeneity (GMM_x) and innocuous neglected heterogeneity (QML_{xv}).

In Fig. 2, we consider the case where the reduced form of the endogenous covariate is misspecified. We generate x_1 using the linearized logit model given in (24) but estimate QML_{xv} and GMM_{xv} using the following linearized loglog reduced form:

$$-\ln\left(-\ln x_1\right) = z\pi_2 + v. \tag{26}$$

The results reported in Fig. 2 show clearly that, under misspecification of the reduced form, the only estimators that produce reliable estimates of structural parameters are GMM_z and LIN_z. All the other estimators are inconsistent when x_1 is endogenous, with the extent of the bias depending on the degree of endogeneity. In effect, the use of an incorrect reduced form is innocuous for the consistency of the GMM_{xv} and LIN_{xv} estimators only when there is no endogeneity, i.e., precisely in the case where no reduced form for x_1 would need to be specified. In terms of RMSE, the GMM_z and LIN_z estimators display the most uniform behavior of all estimators, but there are several cases, particularly for $s = 3$ and $N = 200$, where the RMSE of GMM_z is not among the lowest. However, as the sample size grows, the RMSE of the GMM_z estimator decreases substantially, outperforming most of the other estimators also in terms of this criteria when $N = 1{,}000$. Nevertheless, the RMSE of the LIN_z estimator is always smaller. Finally, note that, with the exception of the GMM_z and LIN_z estimators, the empirical coverage of a 95% confidence interval converges often to zero for all the other estimators. In particular, note the unreliable behavior of the GMM_{xv} estimator: for some values of ρ_{uv}, that coverage converges to zero as N grows; for other values of ρ_{uv}, it seems to converge to one. The same applies, naturally, to the LIN_{xv} estimator.

5.1.3. Conditional partial effects

We now examine the ability of *QML*, *GMM*, and *LIN* estimators to measure partial effects conditional only on observables. For each *GMM* and *LIN* estimator, the partial effect is given by $N^{-1}\hat{\theta}_1 \sum_{j=1}^{N} g\left(\hat{\theta}_0 + \hat{\theta}_1\bar{x}_1 + \hat{u}_j\right)$, where $\bar{x}_1$ represents represents one of the $\{0, 0.02, 0.04, \ldots, 0.98, 1\}$ population quantiles of x_1. To stress the importance of using the proposed two-step procedure for computing conditional partial effects, we estimated also "naive" partial effects, given simply by $\hat{\theta}_1 g\left(\hat{\theta}_0 + \hat{\theta}_1\bar{x}_1\right)$, which sets $u = 0$ in the evaluation of the partial effect. For the *QML* estimators, we computed also naive partial effects and, only for QML_{xv}, the following smearing-type estimator, suggested by Wooldridge (2005): $N^{-1}\hat{\theta}_1 \sum_{j=1}^{N} g\left(\hat{\theta}_0 + \hat{\theta}_1\bar{x}_1 + \hat{\rho}_{uv}\hat{v}_j\right)$. We add the superscript "s" to all estimators that average out the unobservables, e.g., GMM_z^s.

In Fig. 3, we display the mean across the replications of the estimated partial effects for some selected cases. In particular, only three experimental designs are presented in Fig. 3: neglected heterogeneity but no endogeneity ($\rho_{uv} = 0$); strong endogeneity but innocuous neglected heterogeneity ($\rho_{uv} = 1$); and the previous situation but under misspecification of the reduced form. In the three cases, we consider $N = 1{,}000$ and $s = 12$ and display only the partial effects based on the QML_{xv}, GMM_z, and LIN_z estimators. As benchmark, we display also the "true" partial effects, which were calculated by integration as in (3) with $f\left(u|x\right)$ replaced by the density used to generate u.

Under neglected heterogeneity and no endogeneity (first graph), note how both the QML_{xv} and QML_{xv}^s estimators yield partial effects very close to the true ones (the same happens with the not reported QML_x estimator), in spite of being based on inconsistent estimates of the structural parameters. This is in accordance with the conjecture by Wooldridge (2005) that when heterogeneity is independent of the covariates and the interest lies in average partial effects of the observed covariates on mean responses, one may simply ignore the unobservables. Wooldridge (2005) demonstrated this result for the probit model with normal-distributed unobservables, but, given the similarity between probit and logit models, it is not surprising that the same conclusion holds approximately for the specification considered in this Monte Carlo study. Regarding the *LIN* and *GMM* estimators, application of the smearing corrections is clearly essential for estimating consistently conditional partial effects. Otherwise, large biases may be created.

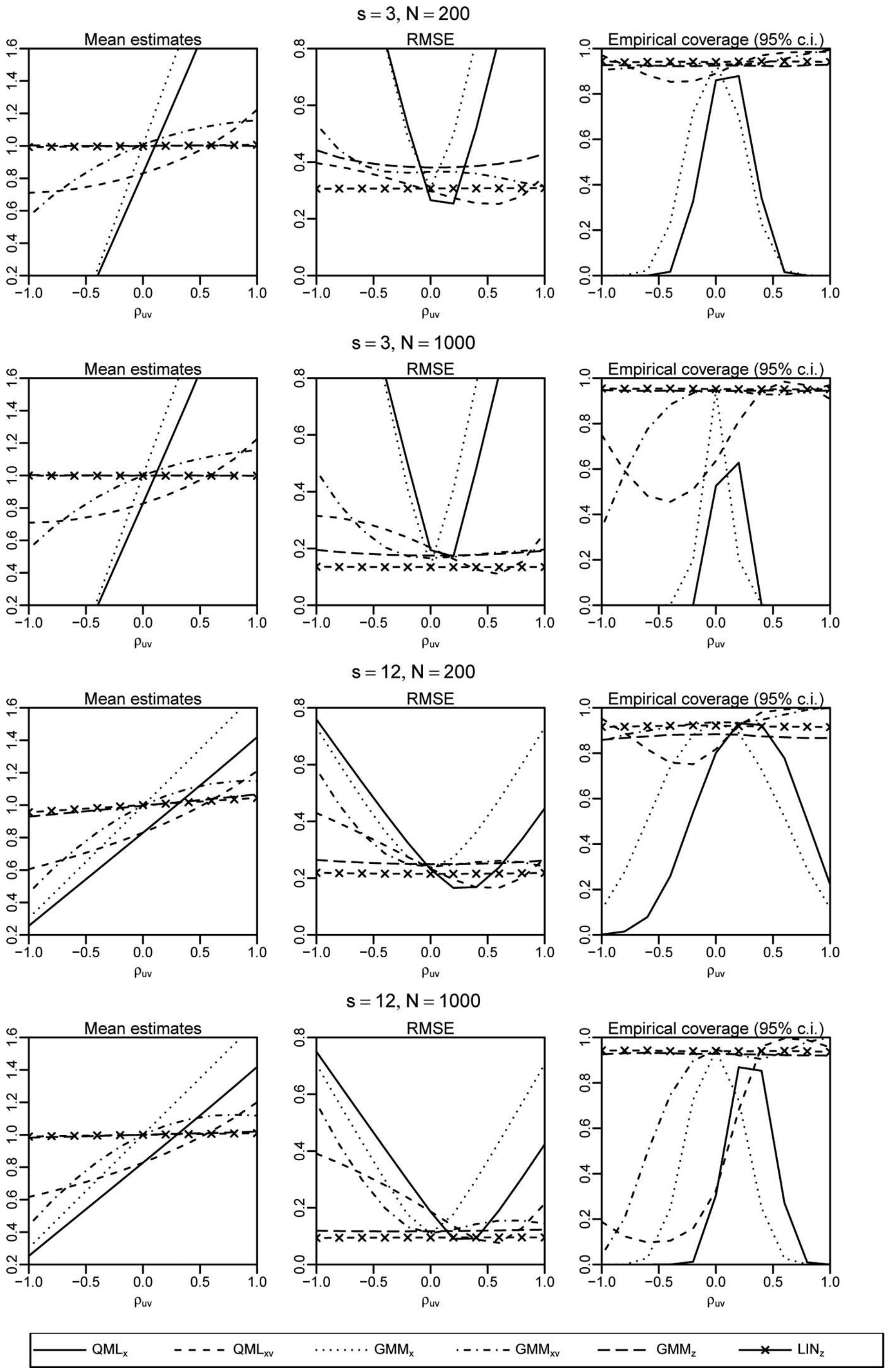

Figure 2. Monte Carlo statistics for the structural parameter θ_1 (incorrect reduced functional form).

When all relevant heterogeneity concerns the endogeneity of x_1 (second and third graphs), all naive estimators provide biased estimates of partial effects. Applying the smearing correction, the GMM_z^s and LIN_z^s estimators are the only ones that estimate consistently the partial effects in both cases. In the case of QML_{xv}^s (and also GMM_{xv}^s and LIN_{xv}^s) it is essential to use the right reduced form for x_1.

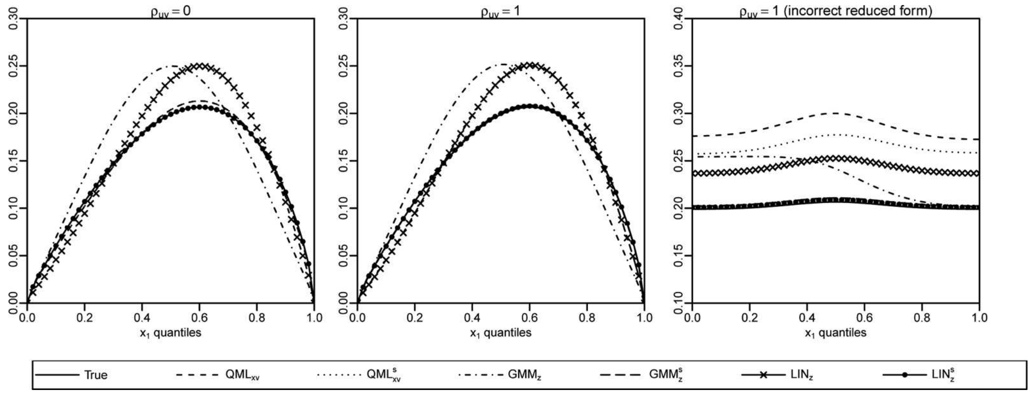

Figure 3. Partial effects conditional on unobservables ($s = 12$; $N = 1,000$).

5.2. Experiments with boundary observations and rounding errors

The results in the previous section revealed a promising behaviour for the *GMM* estimators proposed in this article, but also showed that *LIN* estimators display less variability, especially in small samples. Now, we investigate whether that advantage holds when the sample has boundary observations and, hence, ad-hoc modified *LIN* estimators have to be implemented.

First, we generate the data as in the previous set of experiments, but with the following differences: only (23) is used to generate x_1; the variance of u (σ^2) take values in the interval $\{0.25, 0.5, 1, \ldots, 4\}$; and θ_0 take values in the interval $\{-4, -3.5, \ldots, 0\}$. Then, in order to mimic the rounding errors in official statistics, a new random variable y^* was generated by rounding to the nearest thousandth the values of y obtained in the first stage. This procedure generates a larger number of zeros as θ_0 decreases: for $\sigma^2 = 4$, the average percentage of zeros in the simulated samples ranges from about 3% to about 29% as θ_0 decreases from 0 to -4. It also generates a larger number of zeros as σ^2 increases: for $\theta_0 = -4$, the average percentage of zeros in the simulated samples ranges from about 7% to about 29% as σ^2 increases from 0.25 to 4. As noted by Santos Silva and Tenreyro (2006) in a similar Monte Carlo study for exponential regression models, because the initial model generates a larger proportion of observations close to zero than to one, rounding down is more frequent than rounding up, which will necessarily bias the estimates, since the probability of rounding up or down depends on the covariates.

Under these conditions, the *LIN* estimators cannot be directly applied. Focusing on LIN_z-type estimators, the following two estimators are considered: $LIN_z^{0.001}$, which adds 0.001 to each observation of y^*; and LIN_z^+, which is obtained by dropping the observations for which y^* equals zero. Figure 4 reports the results obtained for GMM_z, $LIN_z^{0.001}$ and LIN_z^+ for some of the experiments performed.

While the *GMM* estimators are relatively robust to the presence of rounding errors, the two *LIN* estimators seem to be very sensitive to this type of measurement error, leading to sizable biases in most cases. Indeed, even for $\left(\theta_0, \sigma^2\right) = (0, 4)$ and $\left(\theta_0, \sigma^2\right) = (-4, 0)$, the cases where the percentages of zeros in the original sample is the lowest, the biases of both estimators are at least 7.5% and 4.6%, respectively. Moreover, their coverage of 95% confidence intervals tends to zero very fast as θ_0 decreases and is nearly zero whenever $\theta_0 = -4$. In contrast, in the former case there is only a slight coverage decrease in the coverage of *GMM* estimator as the percentage of zeros increases, while in the latter case only for $\sigma^2 > 1$ is its empirical coverage below 80%. The performance of the *GMM* estimator is thus clearly encouraging, since, at least in this example, it seems to be able to deal simultaneously with endogeneity, boundary observations, and rounding errors in a more robust way than *LIN* estimators.

5.3. RESET test

The first set of experiments performed in this study was designed in such a way that both *LIN* and *GMM*, based on appropriate instruments, are able to deliver consistent estimators for the slope parameter, since

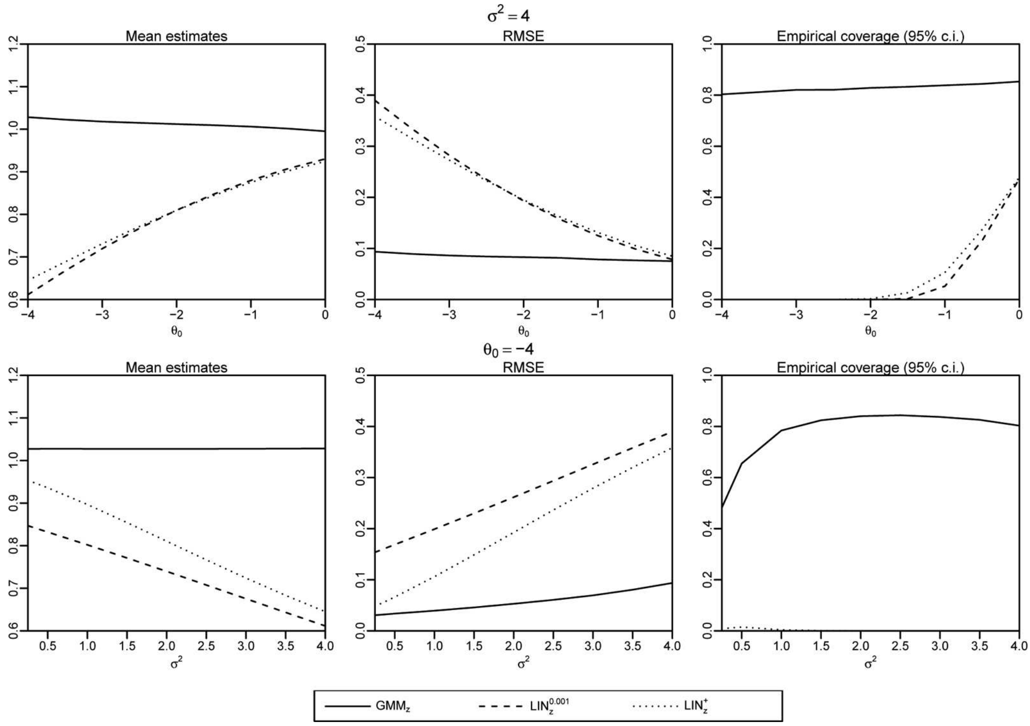

Figure 4. Monte Carlo statistics for the structural parameter with rounding ($s = 12$; $N = 1,000$; $\rho_{uv} = 0.5$).

both $E\left[\exp(u)|z\right]$ and $E(u|z)$ do not depend on z. Now, we consider a data generating process where only one of those two assumptions holds and investigate the ability of the popular RESET test to detect model misspecification. As found out by Ramalho and Ramalho (2012) for binary regression models (which use the same specifications as fractional regression models), the RESET test is sensitive to a large number of model misspecifications, including neglected heterogeneity, and thus it may be useful also in this context.

Equation (22) is again used to generate y, but x_1 and u are now drawn from normal distributions with mean μ and variance σ^2. In the case of x_1, $\mu = 0$ and $\sigma^2 = 0.5$. In the case of u, $\sigma^2 = 1 + \lambda x_1$ and μ is either 0 or $-0.5\,(1 + \lambda x_1)$. In the former case, $E(u|x_1) = 0$ and $E\left[\exp(u)|x_1\right] = \exp\left[0.5\,(1 + \lambda x_1)\right]$, so LIN_x provides consistent estimators for θ_1 irrespective of the value of λ, while GMM_x is consistent only when $\lambda = 0$. In the other case, it occurs the opposite, since $E(u|x_1) = -0.5\,(1 + \lambda x_1)$ and $E\left[\exp(u)|x_1\right] = 1$.

Figure 5 reports the results obtained for an heteroskedasticity-robust Wald version of the RESET test based on the addition of a quadratic power of the index $\left(\hat{\theta}_0 + \hat{\theta}_1 x_1\right)$ to the regression equations underlying the LIN_x and GMM_x estimators, considering $N = \{1000, 5000\}$. The displayed percentage of rejections of the null hypothesis of correct model specification is based on the use of asymptotic critical values and a 5% significance level. Clearly, the RESET test seems to be useful in this context: its power tends to one as λ and/or N increase and its estimated size is either very close to the nominal size ($E(u|x_1) = 0$) or slightly higher but converging to the nominal size as N increases ($E\left[\exp(u)|x_1\right] = 1$).

6. Empirical application: the determinants of corporate capital structure

In this section we use some of the estimators discussed before to assess the determinants of firms' capital structure decisions, namely their option between long-term debt and equity. First, two competing

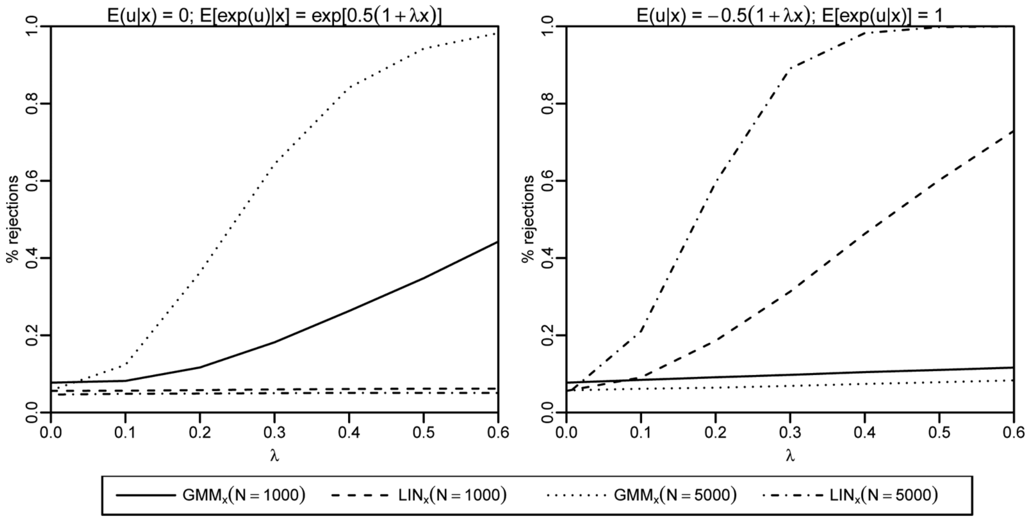

Figure 5. Monte Carlo statistics for the RESET test.

capital structure theories are briefly discussed, then the main characteristics of the data and variables are described, and finally the main estimation results are presented.

6.1. Capital structure theories

Two of the most popular explanations of firms' debt policy decisions are the trade-off and the pecking-order theories. According to the former, firms choose the proportion of debt in their capital structure that maximizes their value, balancing the benefits of debt (e.g., the tax deductibility of interest paid) against its costs (e.g., potential bankruptcy costs caused by an excessive amount of debt). In contrast, the pecking-order theory advocates that, due to information asymmetries between firms' managers and potential outside financiers, firms tend to adopt a perfect hierarchical order of financing, giving preference to the use of internal funds and issuing new shares only when their ability to issue safe debt is exhausted. For details on both theories, see the recent survey by Frank and Goyal (2008).

To evaluate the trade-off and pecking-order theories, many different tests have been proposed in the financial literature. The most common procedure is to use regression models to examine how a given set of potential explanatory variables influences some leverage ratio (*e.g.*, debt to capital or total assets) and then test whether each variable behaves or not as predicted by each theory. Hence, in this framework, the main interest of the econometric analysis lies on the significance of the structural parameters that appear in the leverage equation.

6.2. Data and variables

The data set used in this study was provided by the *Banco de Portugal* Central Balance Sheet Data Office and has already been considered by Ramalho and Silva (2009). It comprises financial information and other characteristics of 4,692 non-financial Portuguese firms for the year 1999. In accordance with the latest definitions adopted by the European Commission (recommendation 2003/361/EC), each firm is assigned to one of the following two size-based group of firms: micro and small firms; medium and large firms. A separate econometric analysis for each group is performed.

As a measure of financial leverage, the ratio of long-term debt (defined as the total company's debt due for repayment beyond one year) to long-term capital assets (defined as the sum of long-term debt and equity) is considered (*Leverage*). In all alternative regression models estimated next, the same explanatory variables as those employed by Ramalho and Silva (2009) are contemplated: *Non-debt tax shields* (*NDTS*), measured by the ratio between depreciation and earnings before interest, taxes and depreciation; *Tangibility*, the proportion of tangible assets and inventories in total assets; *Size*, the natural logarithm of sales; *Profitability*, the ratio between earnings before interest and taxes and total assets; *Growth*, the yearly percentage change in total assets; *Age*, the number of years since the foundation of the firm; *Liquidity*, the sum of cash and marketable securities, divided by current assets; and four activity sector dummies: *Manufacturing*; *Construction*; *Trade* (wholesale and retail); and Transport and *Communication*. Some of these variables are expected to have a positive impact on leverage ratios (e.g., *Profitability* and *Liquidity*, in the case of the trade-off theory; *Growth*, in the case of the pecking-order theory; and *Tangibility* and *Size*, in both cases), while others are expected to have a negative effect (e.g., *NDTS* and *Growth*, in the former theory; and *Profitability*, *Age*, and *Liquidity*, in the latter); see inter alia Ramalho and Silva (2009) for a detailed explanation of these effects.

Table 2 reports descriptive statistics for the dependent and explanatory variables by group of firms. Clearly, the group of medium and large firms display a mean leverage ratio that is substantially higher than that of the other group. While this difference may be partially explained by the variables included in the leverage regression, there are many other factors that may affect the capital structure decisions of firms and that, due to data unavailability, typically are not considered in applied work. For example, especially for smaller firms, it is often argued that the personal characteristics of the firms' owners are important factors for explaining firms' financial leverage decisions; see inter alia Hutchinson (1995). As discussed in previous sections, not accounting for these characteristics may lead to inconsistent estimation of the structural parameters and erroneous conclusions about their significance. As illustrated by the previous example, unobserved heterogeneity may be particularly important for smaller firms. Actually, note that even with respect to the observed variables the smaller firms in our data set are clearly more heterogenous than larger firms: with the exception of *Age*, all other explanatory variables display larger standard deviations for the micro and small group of firms.

6.3. Econometric analysis

Given that leverage ratios are, by definition, bounded on the closed interval [0,1], several authors have recently explained firms' financing decisions using regression models suitable to deal with fractional

Table 2. Descriptive statistics.

Group	Variable	Mean	Min	Max	St.Dev.
Micro and Small	Leverage	0.074	0	0.998	0.189
	NDTS	0.829	0	102.150	3.189
	Tangibility	0.298	0	0.995	0.233
	Size	12.951	7.738	16.920	1.437
	Profitability	0.147	0	1.590	0.117
	Growth	14.923	−81.248	681.354	39.835
	Age	18.267	6	210	12.306
	Liquidity	0.227	0	1	0.247
	Manufacturing	0.563	0	1	0.496
	Construction	0.213	0	1	0.409
	Trade	0.030	0	1	0.171
	Communication	0.116	0	1	0.321
Medium and Large	Leverage	0.148	0	0.978	0.199
	NDTS	0.829	0	26.450	1.479
	Tangibility	0.377	0.002	0.977	0.197
	Size	15.814	11.736	22.270	1.386
	Profitability	0.135	0.001	1.040	0.087
	Growth	8.909	−61.621	188.035	21.014
	Age	28.769	5	184	20.139
	Liquidity	0.120	0	0.963	0.155
	Manufacturing	0.767	0	1	0.423
	Construction	0.121	0	1	0.327
	Trade	0.017	0	1	0.129
	Communication	0.046	0	1	0.210

responses; see inter alia Cook et al. (2008) and Ramalho and Silva (2009). In their formulations, no unobserved heterogeneity is allowed for. Here, we also assume that all observed explanatory variables are exogenous but, in contrast to those authors, allow for the presence of unobservables.

As the minimum value observed for the dependent variable *Leverage* is zero but the maximum is lower than one, see Table 2, GMM_x estimators based on both the fractional logit and complementary loglog models that appear in Table 1 may be used in this context. In contrast, the corresponding LIN_x estimators cannot be directly applied. Next, we restrict our attention to logit-based regression models, considering the following structural model:

$$y = \frac{\exp(x\theta + u)}{1 + \exp(x\theta + u)}. \tag{27}$$

We consider five alternative estimators for (27). The first is the QML_x estimator used by Ramalho and Silva (2009), which ignores the presence of unobservables in (27). The second is the GMM_x estimator, which yields consistent estimators for the structural parameters under the assumption that $E\left[\exp(u)|x\right]$ does not depend on x. The three other estimators are the *OLS* estimators LIN_x^+, which drops all observations with $y = 0$, and $LIN_x^{0.001}$ and $LIN_x^{0.00001}$, which add, respectively, 0.001 and 0.00001 to the value observed for y for all firms.[5] In the case of LIN_x^+, the sample size is reduced by almost 82% (smaller firms) and 50% (larger firms), given the large number of sampled firms that do not use long-term debt. To check the adequacy of each model, we apply the RESET test described in Section 5.3.

Table 3 presents the estimation outcomes resulting from the five techniques. The first point to notice is that the truncation applied to y by the LIN_x^+ estimator originates in several cases very different conclusions from all the other estimators. For instance, the variables *Tangibility* and *Liquidity* for medium and large firms and *Trade* for micro and small firms are important determinants of leverage ratios in all cases except LIN_x^+. Conversely, for the latter group of firms, *Growth* and *Age* are relevant covariates only when the model is estimated by LIN_x^+. Moreover, note how the effect of the variable *Size* differs dramatically between LIN_x^+ and the other estimators: according to LIN_x^+, *Size* affects negatively the proportion of debt used by all firms; according to the other estimators, that effect is positive, as predicted by both the trade-off and pecking-order theories. Clearly, the standard approach in many areas of dropping observations not accommodated by the specified model does not seem to be a recommendable practice in the regression analysis of leverage ratios.

Adding a constant to the value observed for y does not seem to be a good idea either. Indeed, although in terms of parameter significance the conclusions produced by both $LIN_x^{0.001}$ and $LIN_x^{0.00001}$ are identical, in terms of magnitude there are substantial differences. Typically, the regression coefficients of the latter model are more than 1.5 times the parameter estimates of the former, although in some cases they may be also much lower (e.g., the *Construction* coefficient for micro and small firms). Therefore, as the estimates are very sensitive to the value of the constant added, and this has to be defined in an arbitrary way, application of corrections of this type to overcome the problem of boundary observations may often not be a good option.

The results produced by the QML_x and GMM_x estimators are relatively similar in terms of the significance of the parameters. However, particularly for the group of micro and small firms, the same does not happen in terms of the magnitude of the parameters. Moreover, while for the larger group of firms in half of the cases the parameter estimates from QML_x are higher than those from GMM_x and in the other half it happens the opposite, for the group of micro and small firms the regression coefficients are systematically much larger (in absolute value) for GMM_x (the only exception is the variable *Tangibility*). Given that the most typical effect of neglected heterogeneity is the production of an attenuation bias in the estimation of regression coefficients, these results suggest that, as anticipated, neglected heterogeneity may be a very important issue in capital structure studies involving small firms and, hence, that the *GMM* estimators proposed in this paper may be particularly useful in this context. This conjecture is also supported by the RESET test, which in the case of micro and small firms rejects the hypothesis of correct specification of all models except the one that generates the GMM_x estimator.

[5]Note that the highest value that we could add to y is any positive value below 0.002, given that the maximum value for *Leverage* in the sample is 0.998, see Table 2.

Table 3. Logit fractional regression models for capital structure decisions.

	Micro and small firms					Medium and large firms				
	LIN^{+}	$LIN^{0.001}$	$LIN^{0.00001}$	QML_X	GMM_X	LIN^{+}	$LIN^{0.001}$	$LIN^{0.00001}$	QML_X	GMM_X
NDTS	0.0508	−0.0130**	−0.0234**	−0.0316	−0.0961**	−0.0798*	−0.1070***	−0.1726***	−0.1106**	−0.1101
	(0.0368)	(0.0062)	(0.0106)	(0.0314)	(0.0471)	(0.0453)	(0.0352)	(0.0625)	(0.0506)	(0.0769)
Tangibility	0.0241	0.8595***	1.5389***	1.0838***	0.5625	0.3472	3.1608***	5.5869***	1.5296***	1.5626***
	(0.3017)	(0.1984)	(0.3321)	(0.2289)	(0.6183)	(0.2927)	(0.4380)	(0.7531)	(0.2601)	(0.4038)
Size	−0.1035*	0.4523***	0.7818***	0.4757***	0.4861***	−0.1379***	0.3794***	0.7278***	0.1069***	0.1207**
	(0.0535)	(0.0305)	(0.0512)	(0.0379)	(0.1343)	(0.0410)	(0.0600)	(0.1052)	(0.0327)	(0.0570)
Profitability	−3.3667***	−0.9298***	−1.4278***	−2.8171***	−7.3232***	−4.1089***	−5.5156***	−8.8397***	−4.6225***	−6.2095***
	(0.7767)	(0.3108)	(0.5321)	(0.5865)	(1.4259)	(0.6687)	(0.8541)	(1.4932)	(0.6483)	(1.0478)
Growth	0.0039**	−0.0007	−0.0014	−0.0006	−0.0031	0.0073*	0.0133***	0.0212***	0.0069***	0.0121***
	(0.0019)	(0.0009)	(0.0014)	(0.0013)	(0.0030)	(0.0039)	(0.0041)	(0.0067)	(0.0021)	(0.0037)
Age	−0.0145***	0.0018	0.0050	−0.0048	−0.0153	−0.0025	−0.0031	−0.0048	−0.0013	−0.0035
	(0.0050)	(0.0034)	(0.0058)	(0.0042)	(0.0127)	(0.0027)	(0.0039)	(0.0068)	(0.0021)	(0.0030)
Liquidity	−0.1233	−0.5691***	−0.9837***	−1.0443***	−1.8051***	−0.4711	−3.1700***	−5.5614***	−1.8594***	−1.2685**
	(0.3381)	(0.1566)	(0.2640)	(0.3174)	(0.5650)	(0.4273)	(0.4831)	(0.8420)	(0.4023)	(0.6085)
Manufacturing	−0.0427	−0.1374	−0.2240	0.0803	−1.4319*	−0.2382	−0.4247	−0.6476	−0.2709	−0.2322
	(0.2187)	(0.1801)	(0.3004)	(0.1806)	(0.7542)	(0.2091)	(0.3576)	(0.6116)	(0.1917)	(0.2327)
Construction	1.0538***	0.0757	0.0088	0.5332**	-0.7573	0.0918	−0.4349	−0.7943	−0.0598	0.0023
	(0.2708)	(0.2024)	(0.3345)	(0.2103)	(0.8150)	(0.2603)	(0.4444)	(0.7591)	(0.2483)	(0.2987)
Trade	−0.5838	−0.7702***	−1.2566***	−0.7066*	−2.4408***	−0.6743	−1.4412**	−2.5095**	−0.7708	−1.2116**
	(0.4885)	(0.2655)	(0.4508)	(0.3953)	(0.9231)	(0.8856)	(0.6185)	(1.0786)	(0.5415)	(0.5689)
Communication	−0.1954	−0.3083	−0.5255	−0.4514*	−2.5579***	0.1915	−0.5023	−0.9806	−0.0520	0.0484
	(0.2851)	(0.1934)	(0.3261)	(0.2596)	(0.7509)	(0.3155)	(0.5039)	(0.8623)	(0.2635)	(0.3547)
Constant	1.3374	−11.4885***	−19.5024***	−8.7194***	−4.8923***	1.7168**	−9.5426***	−17.2074***	−2.9893***	−2.5685***
	(0.8229)	(0.4349)	(0.7312)	(0.5777)	(1.8807)	(0.7201)	(1.0586)	(1.8557)	(0.5761)	(0.9612)
Number of observations	616	3397	3397	3397	3397	661	1295	1295	1295	1295
RESET p-value	0.0099***	0.0000***	0.0000***	0.0439**	0.2188	0.3637	0.6414	0.7300	0.3267	0.4995

Below the coefficients we report robust standard errors in parentheses;***, ** and * denote coefficients or test statistics which are significant at 1%, 5%, or 10%, respectively.

Table 4. Logit fractional regression models for capital structure decisions - partial effects.

	Micro and small firms					Medium and large firms				
	LIN^{+}	$LIN^{0.001}$	$LIN^{0.00001}$	QML_X	GMM_X	LIN^{+}	$LIN^{0.001}$	$LIN^{0.00001}$	QML_X	GMM_X
Conditional on both observables and unobservables ($u = 0$)										
NDTS	0.0110	−0.0001	0.0000	—	−0.0136	−0.0145	−0.0030	−0.0015	—	−0.0181
Tangibility	0.0052	0.0036	0.0003	—	0.0798	0.0630	0.0900	0.0500	—	0.2569
Size	−0.0224	0.0019	0.0001	—	0.0690	−0.0250	0.0108	0.0065	—	0.0198
Profitability	−0.7278	−0.0039	−0.0002	—	−1.0390	−0.7457	−0.1571	−0.0792	—	−1.0208
Growth	0.0009	0.0000	0.0000	—	−0.0004	0.0013	0.0004	0.0002	—	0.0020
Age	−0.0031	0.0000	0.0000	—	−0.0022	−0.0004	−0.0001	0.0000	—	−0.0006
Liquidity	−0.0266	−0.0024	−0.0002	—	−0.2561	−0.0855	−0.0903	−0.0498	—	−0.2085
Manufacturing	−0.0092	−0.0006	0.0000	—	−0.2032	−0.0432	−0.0121	−0.0058	—	−0.0382
Construction	0.2278	0.0003	0.0000	—	−0.1074	0.0167	−0.0124	−0.0071	—	0.0004
Trade	−0.1262	−0.0033	−0.0002	—	−0.3463	−0.1224	−0.0411	−0.0225	—	−0.1992
Communication	−0.0422	−0.0013	−0.0001	—	−0.3629	0.0347	−0.0143	−0.0088	—	0.0080
Conditional only on observables										
NDTS	0.0091	−0.0004	−0.0005	−0.0020	−0.0022	−0.0134	−0.0076	−0.0094	−0.0135	−0.0086
Tangibility	0.0043	0.0252	0.0350	0.0700	0.0129	0.0585	0.2247	0.3050	0.1862	0.1218
Size	−0.0185	0.0132	0.0178	0.0307	0.0112	−0.0232	0.0270	0.0397	0.0130	0.0094
Profitability	−0.6026	−0.0272	−0.0325	−0.1820	−0.1682	−0.6922	−0.3920	−0.4826	−0.5626	−0.4839
Growth	0.0007	0.0000	0.0000	0.0000	−0.0001	0.0012	0.0009	0.0012	0.0008	0.0009
Age	−0.0026	0.0001	0.0001	−0.0003	−0.0004	−0.0004	−0.0002	−0.0003	−0.0002	−0.0003
Liquidity	−0.0221	−0.0167	−0.0224	−0.0675	−0.0415	−0.0794	−0.2253	−0.3036	−0.2263	−0.0989
Manufacturing	−0.0076	−0.0040	−0.0051	0.0052	−0.0329	−0.0401	−0.0302	−0.0354	−0.0330	−0.0181
Construction	0.1886	0.0022	0.0002	0.0344	−0.0174	0.0155	−0.0309	−0.0434	−0.0073	0.0002
Trade	−0.1045	−0.0225	−0.0286	−0.0456	−0.0561	−0.1136	−0.1024	−0.1370	−0.0938	−0.0944
Communication	−0.0350	−0.0090	−0.0119	−0.0292	−0.0588	0.0323	−0.0357	−0.0535	−0.0063	0.0038

As commented on before, empirical capital structure studies typically focus on the analysis of the significance of structural parameters. Nevertheless, in Table 4, for completeness, we report the estimated partial effects for each model variable. For GMM_x and LIN_x estimators, we report two types of partial effects, one conditional only on observables, using the smearing technique to average out the error term, and the other also conditional on unobservables, setting $u = 0$ (naive partial effects). In the case of QML_x, only the former type of partial effect is calculated, given that this estimator assumes no neglected heterogeneity. In all cases, the values reported are the average sample effects, which are calculated as the mean of the partial effects calculated independently for each firm in the sample.

Table 4 shows clearly that there may be substantial differences between the two types of partial effects, which illustrates the importance of using smearing-type techniques when computing partial effects in nonlinear models with unobservables. It also confirms that neglected heterogeneity seems to affect much more the regression equations estimated for micro and small firms, given that the difference between naive and smearing-corrected partial effects is typically much larger for this group of firms. For example, in the case of GMM_x, for medium and large firms the naive effects are about twice the smearing-corrected effects, while for micro and small firms the former effects are more than six times the latter.

Overall, the results found for the robust GMM_x estimator reinforce the conclusion achieved by Ramalho and Silva (2009) that the pecking-order model provides a better explanation of the capital structure decisions of Portuguese firms than the trade-off theory. Indeed, the effects on leverage of the variables *Tangibility* (+), *Size* (+), *Profitability* (-), *Liquidity* (-), and, in the case of larger firms, *Growth* (+) conform with the former theory and in the last three cases are contrary to the predictions of the latter theory.

7. Conclusion

In this article, we proposed a new transformation regression model to deal with boundary outcomes, neglected heterogeneity and endogeneity issues in a particular class of nonlinear models that treat observed and omitted covariates in a similar manner. The suggested GMM estimators are particularly useful for consistent estimation of structural parameters, since they require only a conditional mean assumption regarding a function of the unobservables. Nevertheless, under some additional assumptions, but still without requiring the full specification of the distribution of the unobservables, our estimators may also be used to estimate partial effects conditional only on observables. One of the estimators proposed has also the very attractive feature of not requiring the specification of a reduced form for the endogenous covariates.

Most of the previous features of the proposed GMM estimator are shared with estimators based on linearized transformations of the structural model. The latter estimators have the obvious advantage of easier implementation, not requiring numerical optimization as GMM. According to the Monte Carlo study undertaken, they also seem to display less variability in small samples than GMM. However, in contrast to the linearized estimators, the proposed GMM estimators have the nice feature of accommodating boundary observations with no need for *ad-hoc* adaptations. While this feature does not immediately imply that GMM estimators should be employed whenever there are boundary values, the example considered in the Monte Carlo study illustrates one situation where their performance is clearly superior to that of linearized estimators. The results obtained in the empirical application also seem to favour the new approach, given the large disparity of results obtained for the three linearized estimators that used different *ad-hoc* transformations to handle the boundary observations. Overall, the proposed GMM approach emerges as an important alternative to the existing linearized estimators, being particularly useful in exponential and fractional regression models with boundary observations.[6]

[6]R code to compute all estimators and tests discussed throughout the article is available at http://evunix.uevora.pt/~jsr/FRM.htm

Acknowledgment

The authors thank the Editor, the Associate Editor, the referees, and Joao Santos Silva for their valuable suggestions and remarks.

Funding

Financial support from Fundacao para a Ciencia e a Tecnologia (grant PTDC/EGE-ECO/119148/2010) is also gratefully acknowledged.

References

Abrevaya, J. (2002). Computing marginal effects in the Box-Cox model. *Econometric Reviews* 21(3):383–393.

Abrevaya, J. (2003). Pairwise-difference rank estimation of the transformation model. *Journal of Business and Economic Statistics* 21(3):437–447.

Abrevaya, J., Hausman, J. A., Khan, S. (2010). Testing for causal effects in a generalized regression model with endogenous regressors. *Econometrica* 78(6):2043–2061.

Cameron, A. C., Trivedi, P. K. (2005). *Microeconometrics. Methods and Applications*. New York: Cambridge University Press.

Cavanagh, C., Sherman, R. P. (1998). Rank estimators for monotic index models. *Journal of Econometrics* 84:351–381.

Chen, S. (2002). Rank estimation of transformation models. *Econometrica* 70(4):1683–1697.

Cook, D. O., Kieschnick, R., McCullough, B. D. (2008). Regression analysis of proportions in Finance with self selection. *Journal of Empirical Finance* 15(5):860–867.

Duan, N. (1983). Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association* 78(383):605–610.

Fève, F., Florens, J. P. (2010). The practice of non-parametric estimation by solving inverse problems: the example of transformation models. *The Econometrics Journal* 13(3):S1–S27.

Frank, M. Z., Goyal, V. K. (2008). "Tradeoff and pecking order theories of debt," In Eckbo B. E. (ed.) *Handbook of Corporate Finance: Empirical Corporate Finance, Vol. 2*, Elsevier, pp. 135–202.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics* 35:303–316.

Heckman, J. J. (2000). Causal parameters and policy analysis in Economics: a twentieth century retrospective. *Quarterly Journal of Economics* 115(1):45–97.

Heckman, J. J. (2001). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy* 109:673–748.

Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* 64(1):103–137.

Horowitz, J. L., Hardle, W. (1996). Direct semiparametric estimation of single-index models with discrete covariates. *Journal of the American Statistical Association* 91(436):1632–1640.

Hutchinson, R. W. (1995). The capital structure and investment decisions of the small owner-managed firm: some exploratory issues. *Small Business Economics* 7:231–239.

Manning, W. G., Mullahy, J. (2001). Estimating log models: to transform or not transform? *Journal of Health Economics* 20:461–494.

Mullahy, J. (1997). Instrumental-variable estimation of count data models: applications to models of cigarette smoking bahavior. *Review of Economics and Statistics* 79(4):586–593.

Newey, W. K., McFadden, D. (1994). "Large sample estimation and hypothesis testing," In Engle, R. F. and McFadden, D. L. (eds.) *Handbook of Econometrics, Vol. 4*. Amesterdam: Elsevier, pp. 2111–2245.

Newey, W. K., West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review* 28:777–787.

Papke, L. E., 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.

Ramalho, J. J. S. (2006). Bootstrap bias-adjusted GMM estimators. *Economics Letters* 92(1):149–155.

Ramalho, E. A., Ramalho, J. J. S. (2010). Is neglected heterogeneity really an issue in binary and fractional regression models? A simulation exercise for logit, probit and loglog models. *Computational Statistics and Data Analysis* 54(4):987–1001.

Ramalho, E. A., Ramalho, J. J. S. (2012). Alternative versions of the RESET test for binary response index models: a comparative study. *Oxford Bulletin of Economics and Statistics* 74(1):107–130.

Ramalho, E. A., Ramalho, J. J. S., Henriques, P. (2010). Fractional regression models for second stage DEA efficiency analyses. *Journal of Productivity Analysis* 34(3):239–255.

Ramalho, E. A., Ramalho J. J. S., Murteira, (2011). Alternative estimating and testing empirical strategies for fractional regression models. *Journal of Economic Surveys* 25(1):19–68.
Ramalho, J. J. S., Silva, J. V. (2009). A two-part fractional regression model for the financial leverage decisions of micro, small, medium and large firms. *Quantitative Finance* 9(5):621–636.
Rivers, D., Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics* 39(3):347–366.
Santos Silva, J. M. C., Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics* 88(4):641–658.
Smith, R. J., Blundell, R. W. (1986). An exogeneity test for a simultaneous equation tobit model with an application to labor supply. *Econometrica* 54(3):679–685.
Vanhems, A., van Keilegom, I. (2011). Semiparametric transformation model with endogeneity: a control function approach. Toulouse School of Economics working-paper 11–243.
Wooldridge, J. M. (1992). Some alternatives to the Box-Cox regression model. *International Economic Review* 33(4):935–955.
Wooldridge, J. M. (1997). "Quasi-likelihood methods for count data," In Pesaran, M. H. and Schmidt, P. (eds.) *Handbook of Applied Econometrics, Vol. II: Microeconomics*, Blackwell, pp. 352–406.
Wooldridge, J. M. (2005). "Unobserved heterogeneity and estimation of average partial effects," In Andrews, D. W. K. and Stock, J. H. (eds.) *Identification and Inference for Econometric Models*, Cambridge University Press, pp. 27–55.
Ye, B. Y. J., Duan, N. (1997). Nonparametric $n^{-1/2}$-consistent estimation for the general transformation models. *Annals of Statistics* 25(6):2682–2717.