Copula-based Tests for Nonclassical Measurement Error – The Case of Fractional Random Variables

JOSÉ M. R. MURTEIRA

CeBER, Faculdade de Economia da Universidade de Coimbra, and CEMAPRE

CeBER Working Papers

No. 13 2018

Copula-based Tests for Nonclassical Measurement Error – The Case of Fractional Random Variables

José M. R. Murteira

CeBER, Faculdade de Economia da Universidade de Coimbra, and CEMAPRE

Abstract

This paper addresses measurement error (ME) of double bounded variables, of which fractional variables, defined on the interval [0,1], constitute a prominent example. The text discusses consequences of ME and suggests a specification test sensitive to ME of such variables. Given the latter's bounded support, ME is not independent of the original error-free variate, a fact that invalidates classical ME assumptions as a framework for the test. This is circumvented with a score test of independence between the error-free variate and ME, under which the latter becomes degenerate at zero and their joint distribution, specified as a copula function, reduces to the original variable's distribution. This procedure yields a specification test of the distribution of the error-free variable, valid under mild assumptions on the marginal distribution of ME and under departures from the specified copula. The test's finite-sample behaviour is also evaluated through a set of simulation experiments.

Keywords:
Copula; Fractional variable; Maximum likelihood; Measurement error; Probability integral transform; Score test.

JEL classification: C12, C25.

1 Introduction

The present paper addresses the possibility of measurement error (ME) of double bounded random variables, a context in which fractional variates, inherently defined on the interval $[0,1]$, constitute a prominent well-studied example (see, *e.g.*, Papke and Wooldridge, 1996; Ferrari and Cribari-Neto, 2004; Ramalho, Ramalho and Murteira, 2011; and the references therein). The presence of ME in this context raises some interesting research issues that challenge conventional estimation and inference approaches. The present text discusses some of these issues and suggests a specification test of the distribution of continuous double bounded variables, presumably sensitive to ME contamination. The proposed test can be used to assess the latter both in a stand-alone variable and in a regression context with double bounded response.

Let Y and V denote, respectively, the error-free continuous variable, defined on the interval $[a, b] \subset \mathbb{R}$, and ME. As the latter is unobservable, the error-free variable Y is also unobservable if erroneously measured. Meanwhile, the fact that the support of Y is bounded makes it difficult to assume that Y and V are independent without *ad hoc* restrictions on the support of Y and/or V. For instance, under additive ME, with $Z \equiv Y + V$ denoting the error-contaminated observable variable, defined on the same support as Y, V cannot be taken as independent of Y as its conditional support, given $Y = y$, is $[a - y, b - y]$. In the terms of the established ME literature, V cannot be considered "strongly classical" (see, *e.g.*, Chen, Hong and Nekipelov, 2011; Schennach, 2013). Moreover, without excluding the boundaries from the support of Y and Z (taking the open interval $]a, b[$), V cannot even be taken as simply mean-dependent of Y (or "weakly classical") because $\mathrm{E}(V|Y = y)$ is necessarily dependent on y, if not elsewhere, at least for $y = a$ or $y = b$. Note that, in order to allow ME at $Y = a$ (respectively, $Y = b$), one must have $\mathrm{E}(V|Y = a) > 0$ (respectively, $\mathrm{E}(V|Y = b) < 0$) – otherwise V would be degenerate at zero in either boundary. In short, $\mathrm{E}(V|Y = y)$ cannot be constant, irrespective of y. As a consequence, the higher order conditional moments of V given $Y = y$ (for instance, the variance) are also, very likely, functionally dependent of y.

One may think of alternative approaches regarding the composition of error-free and ME variates, aiming at independence thereof. Consider, for simplicity, the most usual case, in which Y is fractional, that is, $a = 0$, $b = 1$ (if this is not the case, it can obviously be produced by an affine transformation of Y). One possible approach considers multiplicative ME, with $Z \equiv YV$ (a framework adopted in the case of duration responses by Chesher, Dumangane and Smith, 2002). This case, however, precludes ME at $Y = 0$ and, in addition, independence of Y and V would impose $0 \leq Z \leq Y$ because, necessarily, $0 \leq V \leq l$ for some $l \leq 1$ (otherwise Z could be greater than one). Yet another possibility would be $Z \equiv Y^V$, which avoids the previous difficulty (allowing Z to be greater than Y) but, nevertheless, excludes ME at $\{0,1\}$ as $Z \equiv Y$ in either boundary. A logarithmic transform, $\log Z \equiv V \log Y$ would still preclude ME at $Y = 1$ (besides restricting the support of Y and Z to $]0,1]$).

In view of the above considerations, additive nonclassical ME is assumed in the remainder of the present text. ME is termed "nonclassical" in the sense that it does not comply with classical assumptions regarding independence from the error-free variate (independence from, or mere uncorrelatedness with, Y). Indeed, in this context, either V and Y are statistically dependent or, conversely, V is degenerate at zero. In other words, independence of Y and ME is equivalent to no ME whatsoever. This observation, in turn, suggests an 'indirect' look at the plausibility of dependence between the error-free variate of interest and ME. As described below, one such approach can lead to a feasible specification test of the distribution of Y, expected to be sensitive to the presence of ME of the double bounded variate of interest. Since V is unobserved, a score-type test, involving estimation of the specified model for Y's data generating process (DGP) under the null hypothesis of no ME, seems a natural choice for such procedure.

The starting point of the proposed analysis takes the joint distribution of (Y, V), F, using a bivariate parametric copula formulation. As is well known – see, *e.g.*, Joe (2014) – copula models formalize the dependence structure of component variables, explicitly discerning this structure from each variable's margin. Many common parametric copulas encompass independence as a special case, corresponding to a particular value of their dependence parameter (or a set of values, if the copula involves more than one parameter). Under these values, F collapses to what is usually termed "independence copula", formally expressed as $F_Y F_V$, with F_Y and F_V denoting the marginal distributions of, respectively, Y and V. Under appropriate smoothness of F as function of its dependence parameters in the neighbourhood of the values yielding independence, the corresponding log-density can provide the basis for a score statistic. As detailed below, this approach yields a valid test for ME under mild assumptions on the marginal distribution of ME, as well as under departures from the adopted copula. As previously mentioned, the suggested procedure can be utilized with either a stand-alone variable (in which case the test involves the probability integral transform, $F_Y(Y)$) or a regression double bounded response with covariates $\boldsymbol{X}$ (involving the conditional probability integral transform, given $\boldsymbol{X}$, $F_{Y|\boldsymbol{X}}(Y|\boldsymbol{X})$).

The remainder of the paper is organized as follows. Section 2 completes the notation and sets the general framework, discussing consequences of ME for the distribution of double bounded variables. Section 3 describes the proposed specification test, detailing its variants and asymptotic distribution. Section 4 presents a Monte Carlo study, illustrating the empirical size and power of the proposed test under various designs. Section 5 concludes the paper and suggests future research.

2 Framework

Let Y and V denote, respectively, as previously defined, the unobservable, error-free, double bounded continuous variable of interest, and unobservable continuous ME. Consider, for convenience, that Y is a fractional variable, $y \in [0,1]$, and assume an additive ME model, $Z \equiv Y + V$, with Z observable and defined on the same support as Y. Denote the unknown joint continuous distribution of (Y, V) as F, the support of which can be expressed in general as $\{(y, v) \in \mathbb{R}^2: -y \leq v \leq 1 - y, 0 \leq y \leq 1\}$.

Consider now the expression of F in terms of a parametric copula function. Formally, write the joint distribution of (Y, V) as $C[u_Y, u_V(\boldsymbol{\delta}); \boldsymbol{\delta}] = F(y, v; \boldsymbol{\delta}) = \Pr(Y \leq y, V \leq v; \boldsymbol{\delta})$, where C denotes some continuous bivariate copula, $u_Y \equiv F_Y(y) = \Pr(Y \leq y)$ and $u_V(\boldsymbol{\delta}) \equiv F_V(v; \boldsymbol{\delta}) = \Pr(V \leq v; \boldsymbol{\delta})$. The copula involves one or more dependence parameters, denoted in general by the column d-vector $\boldsymbol{\delta}$ (in addition to those parameters involved solely in the marginal distributions of Y and V, not made explicit). It is important to note at the outset that, in the present double bounded case, the marginal distribution of ME is in some way affected by $\boldsymbol{\delta}$ (contrarily to Y's marginal distribution, which does not involve $\boldsymbol{\delta}$) – at the very least, V becomes degenerate at zero for $\boldsymbol{\delta} = \mathbf{0}$. Hence the explicit notation $u_V(\boldsymbol{\delta})$. In what follows, incidentally, one may simply use u_V instead of $u_V(\boldsymbol{\delta})$.

Assume without loss of generality that independence of Y and V is attained for $\boldsymbol{\delta} = \mathbf{0}$ (under which value, as mentioned, the variable V becomes degenerate at zero). Consider in detail each case.

i. For $\boldsymbol{\delta} \neq \mathbf{0}$ Y is affected by ME, with Y and V continuous dependent random variables. In this case, the joint density of (Y, V) can be produced by differentiating their joint distribution: in general, for $-y \leq v \leq 1 - y,\ 0 \leq y \leq 1$,

$$f(y, v; \boldsymbol{\delta}) \equiv \frac{\partial^2 F(y, v; \boldsymbol{\delta})}{\partial y \partial v} = \frac{\partial^2 C(u_Y, u_V; \boldsymbol{\delta})}{\partial u_Y \partial u_V} \frac{du_Y}{dy} \frac{du_V}{dv} \equiv$$
$$c(u_Y, u_V; \boldsymbol{\delta}) f_Y(y) f_V(v; \boldsymbol{\delta}), \tag{1}$$
$$\boldsymbol{\delta} \neq \mathbf{0},$$

where $c(u_Y, u_V; \boldsymbol{\delta}) \equiv \partial^2 C(u_Y, u_V; \boldsymbol{\delta}) / \partial u_Y \partial u_V$ and f_Y and f_V denote, respectively, the marginal densities of Y and V. The density of Y may be taken as conditional on a vector of observed covariates, $\boldsymbol{X}$, but this dependence is not made explicit at present.

ii. For $\boldsymbol{\delta} = \mathbf{0}$, Y and V are independent and V is degenerate at zero (no ME) – with Y continuous, unaffected by the value of $\boldsymbol{\delta}$. Thus,

$$u_V(\mathbf{0}) = F_V(v; \mathbf{0}) = \Pr(V \leq v; \mathbf{0}) = \begin{cases} 0, & v < 0 \\ 1, & v \geq 0 \end{cases}, \tag{2}$$

hence, for $0 \leq y \leq 1$,

$$F(y, v; \mathbf{0}) = \Pr(Y \leq y) \Pr(V \leq v; \mathbf{0}) = \begin{cases} 0, & v < 0 \\ F_Y(y), & v \geq 0 \end{cases}.$$

Given that $f_Y(y) = dF_Y(y)/dy$ and

$$f_V(v; \mathbf{0}) = \Pr(V = v; \mathbf{0}) = \begin{cases} 0, & v \neq 0 \\ 1, & v = 0 \end{cases},$$

the joint density, for $0 \leq y \leq 1$, can be written as

$$f(y, v; \mathbf{0}) = \begin{cases} 0, & v \neq 0 \\ f_Y(y), & v = 0 \end{cases}. \tag{3}$$

The next step considers the approximation to the joint density $f(y, v; \boldsymbol{\delta})$ through its Taylor-expansion in the neighbourhood of $\boldsymbol{\delta} = \mathbf{0}$. Formally,

$$f(y, v; \boldsymbol{\delta}) = f(y, v; \mathbf{0}) + \left(\frac{\partial f}{\partial \boldsymbol{\delta}'}\right)_{\boldsymbol{\delta}=\mathbf{0}} \boldsymbol{\delta} + o(\boldsymbol{\delta}) \approx f_Y(y) + \left(\frac{\partial f}{\partial \boldsymbol{\delta}'}\right)_{\boldsymbol{\delta}=\mathbf{0}} \boldsymbol{\delta}, \tag{4}$$

where A_B stands for evaluation of A under condition B, $o(\boldsymbol{\delta})$ denotes the remainder of the expansion, vanishing as $\boldsymbol{\delta} \to \mathbf{0}$ (for which it is sufficient that f have continuous partial derivatives with respect to $\boldsymbol{\delta}$ up to order two in a neighbourhood of $\boldsymbol{\delta} = \mathbf{0}$) and the last expression results from (3). The second term in this expression, involving the gradient of f with respect to $\boldsymbol{\delta}$ evaluated at $\boldsymbol{\delta} = \mathbf{0}$ – $(\partial f / \partial \boldsymbol{\delta}')_{\boldsymbol{\delta}=\mathbf{0}}$ – requires some care as f is defined by different expressions under $\boldsymbol{\delta} \neq \mathbf{0}$ and $\boldsymbol{\delta} = \mathbf{0}$. Let $\boldsymbol{\delta}_{(j)}$ denote a column d-vector with all elements zero except the j-th (denoted as δ_j, for $1 \leq j \leq d$). From (3) the j-th component of $(\partial f / \partial \boldsymbol{\delta})_{\boldsymbol{\delta}=\mathbf{0}}$ can be written as

$$\left(\frac{\partial f}{\partial \delta_j}\right)_{\boldsymbol{\delta}=\mathbf{0}} = \lim_{\delta_j \to 0} \frac{f\big(y, v; \boldsymbol{\delta}_{(j)}\big) - f(y, v; \mathbf{0})}{\delta_j} = \lim_{\delta_j \to 0} \frac{f\big(y, v; \boldsymbol{\delta}_{(j)}\big) - f_Y(y)}{\delta_j}.$$

Applying L'Hôpital's rule, using (1) the limit can be written

$$\lim_{\delta_j \to 0} \frac{c\big[u_Y, u_V\big(\boldsymbol{\delta}_{(j)}\big); \boldsymbol{\delta}_{(j)}\big] f_Y(y) f_V\big(v; \boldsymbol{\delta}_{(j)}\big) - f_Y(y)}{\delta_j} =$$

$$f_Y(y) \lim_{\delta_j \to 0} \frac{\partial \big\{c\big[u_Y, u_V\big(\boldsymbol{\delta}_{(j)}\big); \boldsymbol{\delta}_{(j)}\big] f_V\big(v; \boldsymbol{\delta}_{(j)}\big)\big\}}{\partial \delta_j}, \tag{5}$$

valid on assumption that $c\big[u_Y, u_V\big(\boldsymbol{\delta}_{(j)}\big); \boldsymbol{\delta}_{(j)}\big] f_V\big(v; \boldsymbol{\delta}_{(j)}\big)$ be differentiable with respect to δ_j on an open interval containing zero, except possibly at zero. The derivative can be written as

$$\frac{\partial \big\{c\big[u_Y, u_V\big(\boldsymbol{\delta}_{(j)}\big); \boldsymbol{\delta}_{(j)}\big] f_V\big(v; \boldsymbol{\delta}_{(j)}\big)\big\}}{\partial \delta_j} =$$

$$\left[\frac{\partial c}{\partial u_V\big(\boldsymbol{\delta}_{(j)}\big)} \frac{\partial u_V\big(\boldsymbol{\delta}_{(j)}\big)}{\partial \delta_j} + \frac{\partial c}{\partial \delta_j}\right] f_V\big(v; \boldsymbol{\delta}_{(j)}\big) + c \frac{\partial f_V\big(v; \boldsymbol{\delta}_{(j)}\big)}{\partial \delta_j},$$

where $c \equiv c\big[u_Y, u_V\big(\boldsymbol{\delta}_{(j)}\big); \boldsymbol{\delta}_{(j)}\big]$.

At this point it is necessary to specify in some way the manner in which the marginal distribution of ME is affected by the dependence parameter. Assume that the marginal distribution of ME, $u_V(\boldsymbol{\delta})$, is a differentiable function of its variance, σ_V^2, that the dependence parameter, $\boldsymbol{\delta}$, only affects $u_V(\boldsymbol{\delta})$ through the variance, $\sigma_V^2 \equiv \sigma_V^2(\boldsymbol{\delta})$, and that this variance is a smooth continuous function of $\boldsymbol{\delta}$, such that

$$\begin{cases} \sigma_V^2(\mathbf{0}) = \lim_{\boldsymbol{\delta} \to \mathbf{0}} \sigma_V^2(\boldsymbol{\delta}) = 0 \\ \lim_{\boldsymbol{\delta} \to \mathbf{0}} \dfrac{\partial \sigma_V^2(\boldsymbol{\delta})}{\partial \delta_j} = 0, \forall j \end{cases}. \tag{6}$$

The first part of assumption (6) formalizes the intuition that V is degenerate under $\boldsymbol{\delta} = \mathbf{0}$; given that $Z \equiv Y + V$ is supported on [0,1] for all $\boldsymbol{\delta}$, this degeneracy of V with $\boldsymbol{\delta} = \mathbf{0}$ necessarily occurs at zero. Thus one can write

$$\frac{\partial\{c[u_Y, u_V(\boldsymbol{\delta}_{(j)}); \boldsymbol{\delta}_{(j)}] f_V(v; \boldsymbol{\delta}_{(j)})\}}{\partial \delta_j} =$$

$$\left\{\left[\frac{\partial c}{\partial u_V(\boldsymbol{\delta}_{(j)})} \frac{\partial u_V(\boldsymbol{\delta}_{(j)})}{\partial \sigma_V^2(\boldsymbol{\delta})} \frac{\partial \sigma_V^2(\boldsymbol{\delta})}{\partial \delta_j} + \frac{\partial c}{\partial \delta_j}\right] f_V(v; \boldsymbol{\delta}_{(j)}) + c \frac{\partial f_V(v; \boldsymbol{\delta}_{(j)})}{\partial \sigma_V^2(\boldsymbol{\delta})} \frac{\partial \sigma_V^2(\boldsymbol{\delta})}{\partial \delta_j}\right\},$$

from which

$$\lim_{\delta_j \to 0} \frac{\partial c}{\partial \delta_j} f_V(v; \boldsymbol{\delta}_{(j)}) = \left(\frac{\partial c}{\partial \delta_j}\right)_{\boldsymbol{\delta}=\mathbf{0}} f_V(v; \mathbf{0}) = \left(\frac{\partial c}{\partial \delta_j}\right)_{\boldsymbol{\delta}=\mathbf{0}}, \tag{7}$$

in view of (3). Introducing this result in (5), one can finally write the first-order approximation to the joint density of (Y, V), from (4), as

$$f(y, v; \boldsymbol{\delta}) \approx f_Y(y)[1 + \boldsymbol{D}[F_Y(y)]'\boldsymbol{\delta}], \quad -y \leq v \leq 1 - y, \; 0 \leq y \leq 1, \tag{8}$$

with $\boldsymbol{D}[F_Y(y)]$ denoting the column d-vector

$$\boldsymbol{D}[F_Y(y)] \equiv \left(\frac{\partial c}{\partial \boldsymbol{\delta}}\right)_{\boldsymbol{\delta}=\mathbf{0}}, \tag{9}$$

where $\partial c/\partial \boldsymbol{\delta}$ now represents the vector of 'direct' derivatives of $c[u_Y, u_V(\boldsymbol{\delta}); \boldsymbol{\delta}]$ with respect to $\boldsymbol{\delta}$ (that is, without consideration of the dependence of $u_V(\boldsymbol{\delta})$ on $\boldsymbol{\delta}$) and the expression of $u_V(\mathbf{0})$ involved in $(\partial c/\partial \boldsymbol{\delta})_{\boldsymbol{\delta}=\mathbf{0}}$ is given by (2). If the copula function involves one single-parameter (δ scalar, $d = 1$), (8) can be written as

$$f(y, v; \delta) \approx f_Y(y)\big[1 + \delta D[F_Y(y)]\big], \quad -y \leq v \leq 1 - y, \; 0 \leq y \leq 1. \tag{10}$$

Some remarks with respect to the approximation to $f(y, v; \boldsymbol{\delta})$ seem useful at this point. Firstly, (8) does not explicitly involve the functional form of the distribution of ME – although it obviously depends upon the adopted assumptions on $u_V(\boldsymbol{\delta})$, namely (6). Nonetheless, it does involve the distribution of the error-free variate – through its density, f_Y, and distribution, U_Y. Secondly, for the approximation to provide a useful model for statistical analysis of error-contaminated data, one must ensure that (8) constitutes a proper density, integrating to one and nonnegative over the support of (Y, V).

The first of the above requirements – integration to unity – can be checked to hold, as follows. Consider again (7) and rewrite (8) as

$$f(y, v; \boldsymbol{\delta}) \approx f_Y(y) + \left[\left(\frac{\partial c}{\partial \boldsymbol{\delta}'}\right) f_Y(y) f_V(v; \boldsymbol{\delta})\right]_{\boldsymbol{\delta}=\mathbf{0}} \boldsymbol{\delta},$$

which, under assumption (6), can be written as

$$f_Y(y) + \left(\frac{\partial}{\partial \boldsymbol{\delta}'}\{c[u_Y, u_V(\boldsymbol{\delta}); \boldsymbol{\delta}] f_Y(y) f_V(v; \boldsymbol{\delta})\}\right)_{\boldsymbol{\delta}=\mathbf{0}} \boldsymbol{\delta}.$$

Now, taking the integral over the support of (Y, V), assuming that the order of integration and differentiation can be interchanged, and in view of (1),

$$\int_0^1 \int_{-y}^{1-y} \left\{ f_Y(y) + \left[\left(\frac{\partial}{\partial \boldsymbol{\delta}'} \{c[u_Y, u_V(\boldsymbol{\delta}); \boldsymbol{\delta}] f_Y(y) f_V(v; \boldsymbol{\delta})\} \right)_{\boldsymbol{\delta}=\mathbf{0}} \right] \boldsymbol{\delta} \right\} dv dy =$$

$$\int_0^1 \int_{-y}^{1-y} f_Y(y) dv dy + \frac{\partial}{\partial \boldsymbol{\delta}'} \left\{ \int_0^1 \int_{-y}^{1-y} c[u_Y, u_V(\boldsymbol{\delta}); \boldsymbol{\delta}] f_Y(y) f_V(v; \boldsymbol{\delta}) dv dy \right\}_{\boldsymbol{\delta}=\mathbf{0}} \boldsymbol{\delta} =$$

$$1 + \frac{\partial}{\partial \boldsymbol{\delta}'} \left[\int_0^1 \int_{-y}^{1-y} f(y, v; \boldsymbol{\delta}) dv dy \right]_{\boldsymbol{\delta}=\mathbf{0}} \boldsymbol{\delta} = 1 + \frac{\partial}{\partial \boldsymbol{\delta}'} (1) \boldsymbol{\delta} = 1 + \mathbf{0}' \boldsymbol{\delta} = 1$$

($\mathbf{0}'$ denotes a row d-vector of zeros).

The second requirement – nonnegativity of (8) over the support of (Y, V) or, equivalently, $(\partial c / \partial \boldsymbol{\delta}')_{\boldsymbol{\delta}=\mathbf{0}} \boldsymbol{\delta} \geq -1$ in a neighbourhood of $\boldsymbol{\delta} = \mathbf{0}$ – is satisfied by any continuously differentiable copula function in a small enough neighbourhood of $\boldsymbol{\delta} = \mathbf{0}$ (the case with the examples considered in Section 4).

The first-order expansion (8) thus yields a proper density which provides an approximation to $f(y, v; \boldsymbol{\delta})$ involving the distribution of the error-free variate and the derivative of the copula with respect to its dependence parameter. Given the definition $Z \equiv Y + V$, a standard change-of-variables technique yields the joint density of (Z, V), $f_{ZV}(z, v; \boldsymbol{\delta}) = f(z - v, v; \boldsymbol{\delta})$ with bivariate support $z - 1 \leq v \leq z$, $0 \leq z \leq 1$. Taking (8), an approximation to f_{ZV} can be expressed as

$$f_{ZV}(z, v; \boldsymbol{\delta}) \approx f_Y(z)\{1 + \boldsymbol{D}[F_Y(z)]' \boldsymbol{\delta}\}, \quad z - 1 \leq v \leq z, \quad 0 \leq z \leq 1,$$

where $\boldsymbol{D}[F_Y(z)]$ is similar to $\boldsymbol{D}[F_Y(y)]$ – check (9) – but with u_Y now denoting the evaluation of F_Y at z that is, $F_Y(z)$. From this approximation to f_{ZV} an approximate margin of Z immediately results as

$$f_Z(z; \boldsymbol{\delta}) = \int_{z-1}^{z} f_{ZV}(z, v; \boldsymbol{\delta}) dv \approx f_Y(z)\{1 + \boldsymbol{D}[F_Y(z)]' \boldsymbol{\delta}\} \int_{z-1}^{z} dv =$$

$$f_Y(z)\{1 + \boldsymbol{D}[F_Y(z)]' \boldsymbol{\delta}\}, \quad 0 \leq z \leq 1. \tag{11}$$

With no ME of Y ($\boldsymbol{\delta} = \mathbf{0}$) this expression coincides with $f_Y(y)$ – quite expectably, as $Z \equiv Y$. With covariates $\boldsymbol{X}$ present, the conditional distribution and density of Y given $\boldsymbol{X} = \boldsymbol{x}$ should be used in the expression of the approximate $f_{Z|\boldsymbol{X}}$.

By specifying F_Y (or $F_{Y|\boldsymbol{X}}$) and C, the previous result may be useful if one wishes to estimate parameters while making allowance for the impact of ME. For instance, using a one-parameter copula, if under $F_{Y|\boldsymbol{X}}$ one has $\mathrm{E}(Y|\boldsymbol{x}) = G(\boldsymbol{x}\boldsymbol{\beta})$ – *e.g.*, logit – but ME is suspected to affect the response data, one may want to estimate $\boldsymbol{\beta}$ by using the approximate model $\mathrm{E}(Z|\boldsymbol{x}) = G(\boldsymbol{x\beta}) + \delta\mathrm{E}\{YD[F_{Y|\boldsymbol{X}}(Y|\boldsymbol{x})]|\boldsymbol{x}\}$. Estimating the parameters of this model, the significance of δ can then be tested, in which case acceptance of the null indicates no evidence of ME in the response. Typically, the practical difficulty here concerns the estimation of $\mathrm{E}\{Y \times D[F_{Y|\boldsymbol{X}}(Y|\boldsymbol{x})]|\boldsymbol{x}\}$, possibly requiring some approximation technique.

A general consequence of the foregoing discussion is that the presence of ME of a double bounded variable (as a stand-alone variate or as a regression response) induces distributional changes that lead to inconsistency of the usual estimators of relevant features of the misspecified distribution (including, *e.g.*, parameters of the response's conditional mean given regressors). One such example, frequently applied to models for fractional data, is provided by the Bernoulli-based quasi-maximum likelihood (QML) estimator, which, as is well known, only requires correct specification of $\mathrm{E}(Y|\boldsymbol{x})$ – see, *e.g.*, Papke, *et al.* (1996). In the present context, however, not even $\mathrm{E}(Y|\boldsymbol{x})$ is immune to ME – contrarily, for instance, to the effect of classical additive ME in unbounded responses of linear models (where mismeasurement does not lead to inconsistency but only to less statistical precision in estimation – see, *e.g.*, Hausman, 2001). Consequently, also QML – like nonlinear least squares or maximum likelihood (ML) based on $f_{Y|\boldsymbol{X}}$ alone – will generally prove inconsistent in this case.

The previous arguments suggest that it is prudent to subject the adopted specification of the bounded variate of interest, f_Y or $f_{Y|\boldsymbol{X}}$, to a specification test sensitive to the presence of ME. As already mentioned, one might think of estimating (11) and testing the null hypothesis $H_0: \boldsymbol{\delta} = \mathbf{0}$. However, such procedure would require specification of the unknown copula, C (naturally based, if anything, on analytical and/or computational convenience), so a score test of the null hypothesis of no ME, avoiding estimation of the alternative model, appears as a natural choice. This procedure is described in the next section. Actually, as verified below, the proposed score test leads to a valid procedure, irrespective of the form of the unknown copula function.

3 A Score Specification Test

Employing expression (11) as the basis for a likelihood function of the parameters of f_Z, the score contribution for $\boldsymbol{\delta}$ under the null hypothesis $H_0\colon \boldsymbol{\delta} = \mathbf{0}$ is simply $\boldsymbol{D}[F_Y(z)]$ (again, the distribution of Y can be conditional on covariates). Suppose that F_Y involves the k-vector of parameters $\boldsymbol{\theta}$ and denote this as $F_Y(\cdot;\boldsymbol{\theta})$. Assume further that a random sample of n realisations on the observable fractional variate Z is available, and let $\widehat{\boldsymbol{D}}_{\boldsymbol{n}} \equiv n^{-1/2}\sum_{i=1}^{n}\boldsymbol{D}\left[F_Y\left(z_i;\widehat{\boldsymbol{\theta}}\right)\right]$ (here and elsewhere $\widehat{(\cdot)}$ denotes evaluation at ML estimates of the unknown parameters of F_Y – that is, ML estimates under the null hypothesis of no ME). Then, the score test statistic for $H_0\colon \boldsymbol{\delta} = \mathbf{0}$ is

$$M_n = \widehat{\boldsymbol{D}}_{\boldsymbol{n}}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{-\mathbf{1}}\widehat{\boldsymbol{D}}_{\boldsymbol{n}}, \tag{12}$$

with $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}$ an estimate of $\boldsymbol{\Sigma}$, the asymptotic $d \times d$ covariance matrix of $\widehat{\boldsymbol{D}}_{\boldsymbol{n}}$ under $\boldsymbol{\delta} = \mathbf{0}$. Assuming that the copula and the distribution of the continuous variable Y are correctly specified, under H_0 and usual regularity conditions, M_n is asymptotically distributed as a chi-squared random variate with d degrees of freedom (χ_d^2).

A closer look at the asymptotic behaviour of $\widehat{\boldsymbol{D}}_{\boldsymbol{n}}$ under the null and alternative hypotheses seems useful at this point. Under the null hypothesis ($\boldsymbol{\delta} = \mathbf{0}$ – no ME, $Z \equiv Y$) a random sample of realisations of Y is available so, by a Law of Large Numbers, $\mathrm{plim}_{n\to\infty} n^{-1/2}\widehat{\boldsymbol{D}}_{\boldsymbol{n}} = \mathrm{E}\{\boldsymbol{D}[F_Y(Y)]\}$. Let C denote the adopted copula (which may, or may not, coincide with the DGP copula) with dependence parameter $\boldsymbol{\delta}$. Then, assuming that the order of differentiation and integration can be reversed,

$$\mathrm{E}\{\boldsymbol{D}[F_Y(Y)]\} = \mathrm{E}\left[\left(\frac{\partial c}{\partial \boldsymbol{\delta}}\right)_{\boldsymbol{\delta}=\mathbf{0}}\right] = \left[\frac{\partial}{\partial \boldsymbol{\delta}}\mathrm{E}(c)\right]_{\boldsymbol{\delta}=\mathbf{0}} = \left[\frac{\partial}{\partial \boldsymbol{\delta}}(1)\right]_{\boldsymbol{\delta}=\mathbf{0}} = \mathbf{0}, \tag{13}$$

where the third equality results from the fact that C is itself a proper joint distribution (regardless of whether it is correctly specified or not) so its associated density, c, has expectation one.

One consequence of the previous result, relevant from a practical perspective, is the fact that, in general, any proper bivariate continuous copula (encompassing independence) can be employed to carry out the test – as $\mathrm{plim}_{n\to\infty} n^{-1/2}\widehat{\boldsymbol{D}}_{\boldsymbol{n}} = \mathbf{0}$ under the null, irrespective of the DGP copula (in case of ME contamination). For instance, namely for the sake of simplicity, one can make use of one of the several single-parameter $(d = 1)$ bivariate copulas available in the literature (several examples of which are used in a Monte Carlo study, in Section 4).

If the variable of interest is affected by ME, result (13) is not attained, in general. Note that in this case the available sample comes from $Z \equiv Y + V$, which is not distributed according to F_Y. Consequently, the random variable $W \equiv F_Y(Z)$ is not a probability integral transform, not distributed according to a standard uniform law, $\mathcal{U}(0,1)$, but with density

$$f_W(w; \boldsymbol{\delta}) = \frac{f_Z[F_Y^{-1}(w); \boldsymbol{\delta}]}{f_Y[F_Y^{-1}(w)]} = \frac{f_Z(z; \boldsymbol{\delta})}{f_Y(z)} \approx 1 + \boldsymbol{D}[F_Y(z)]'\boldsymbol{\delta}, \qquad 0 \le w \le 1, \quad (14)$$

where the approximation results from replacing $f_Z(z; \boldsymbol{\delta})$ with expression (11) (as expected, $\boldsymbol{\delta} = \boldsymbol{0}$ yields a uniform density). Therefore, C, as a function of $F_Y(z)$ and $u_V(\boldsymbol{\delta})$, is no longer a proper copula (because it does not have $\mathcal{U}(0,1)$ marginals) or, equivalently, $c[F_Y(z), u_V(\boldsymbol{\delta}); \boldsymbol{\delta}]$ is not a proper bivariate density. Therefore, with ME, $\mathrm{E}\{\boldsymbol{D}[F_Y(Z)]\} \neq \boldsymbol{0}$ and $\widehat{\boldsymbol{D}}_{\boldsymbol{n}}$ is $O_p(\sqrt{n})$ so the test is consistent.

The proposed specification test can be viewed as a test of the d moment conditions $\mathrm{E}\{\boldsymbol{D}[F_Y(Z)]\} = \boldsymbol{0}$ (or $\mathrm{E}\{\boldsymbol{D}[F_{Y|\boldsymbol{X}}(Z|\boldsymbol{X})]|\boldsymbol{X}\} = \boldsymbol{0}$, with covariates $\boldsymbol{X}$). In accordance with Newey (1985) (see also Pagan and Vella, 1989), the expression of the asymptotic covariance matrix of $\widehat{\boldsymbol{D}}_{\boldsymbol{n}}$ under the null hypothesis is generally given by $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{n}} = \boldsymbol{HJH}'$, with

$$\underset{(d+k)\times(d+k)}{\boldsymbol{J}} \equiv n^{-1}\sum_i \begin{bmatrix} \underset{d\times d}{\boldsymbol{D}[F_Y(z_i;\widehat{\boldsymbol{\theta}})]\boldsymbol{D}[F_Y(z_i;\widehat{\boldsymbol{\theta}})]'} & \underset{d\times k}{\boldsymbol{D}[F_Y(z_i;\widehat{\boldsymbol{\theta}})]\hat{\boldsymbol{s}}_{\boldsymbol{i}}'} \\ \underset{k\times d}{\hat{\boldsymbol{s}}_{\boldsymbol{i}}\boldsymbol{D}[F_Y(z_i;\widehat{\boldsymbol{\theta}})]'} & \underset{k\times k}{\hat{\boldsymbol{s}}_{\boldsymbol{i}}\hat{\boldsymbol{s}}_{\boldsymbol{i}}'} \end{bmatrix}$$

$$\underset{d\times(d+k)}{\boldsymbol{H}} \equiv \left[\underset{d\times d}{\boldsymbol{I_d}} \; \vdots \; -\left\{ \underset{d\times k}{n^{-1}\sum_i \frac{\partial \boldsymbol{D}[F_Y(z_i;\widehat{\boldsymbol{\theta}})]}{\partial \boldsymbol{\theta}'}} \right\} \left(\underset{k\times k}{n^{-1}\sum_i \frac{\partial \hat{\boldsymbol{s}}_{\boldsymbol{i}}}{\partial \boldsymbol{\theta}'}} \right)^{-1} \right],$$

where $\boldsymbol{I_d}$ denotes the identity matrix of order d, summations are over $i = 1, \dots, n$ and $\hat{\boldsymbol{s}}_{\boldsymbol{i}} \equiv \partial \log f_Y(z_i; \widehat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta}$ denotes the i-th contribution to the score vector under H_0. If the adopted copula only involves one dependence parameter (δ scalar – $d = 1$), $\boldsymbol{H}$ is a row $(k+1)$-vector and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}$ is a scalar.

An "outer product of the gradient" (OPG) variant of the test, asymptotically equivalent to it, is rather easier to compute as it does not require the construction of the previous matrices. This version of the test is computed as n times the uncentered R^2 from the auxiliary regression

$$1 = \boldsymbol{D}[F_Y(z_i;\widehat{\boldsymbol{\theta}})]'\boldsymbol{\zeta} + \hat{\boldsymbol{s}}_{\boldsymbol{i}}'\boldsymbol{\eta} + \text{error}, \qquad (15)$$

with $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ denoting parameter vectors.

Although much easier to implement, the OPG version of the test is well-known for its poor small sample performance in some leading cases. In the next section, a brief Monte Carlo experiment presents simulation results that illustrate the small sample performance of both versions of the test just described.

4 Monte Carlo Experiment

4.1 General Design

This section presents a Monte Carlo experiment designed to illustrate the finite sample performance of the test described in the paper. Although the test can obviously be applied to any continuous double bounded variable, only the fractional case is considered in the experiment, as this is, by far, the most frequent. The study is organized in two main parts: the first considers ME of a stand-alone fractional variable whereas, in the second part, ME is supposed to affect a fractional response in a regression model. In both parts the two versions of the proposed test are implemented with several single-parameter ($d = 1$) copulas, yielding particular expressions for $D(\cdot)$ and corresponding test statistics. The two versions of the test examined in the experiment are obtained from (12) and (15), named respectively as M1 and M2. Under the various DGP's considered, rejection rates of H_0 are computed at the 5% nominal size (with respect to the χ_1^2 asymptotic null law), based on 10000 random samples of size $n = 500$.(1) All computations were performed using the R software.

In all the alternative DGP's considered, the error-free variate, Y, is a continuous fractional variable distributed according to a Beta DGP (with different parameter values, or different conditional means in the regression case). Under ME (denoted V), the latter is randomly drawn from one of the following continuous conditional distributions, given $Y = y$: $V + y|y \sim Beta(1/(1-y), 2)$ (yielding a conditional mode of the observable fractional variable, Z, equal to y); $V|y$ truncated Normal with

(1) Samples of size 250 and 1000 produce expectable results, in line with those reported in the paper, and are therefore omitted. Rejection rates obtained at 1% and 10% nominal levels are also not reported, as they meet overall expectations, given reported rates at the 5% nominal level.

parameters (0,1) and support $[-y/2\,,(1-y)/2]$ – $\mathcal{NT}(0,1)$; $V|y$ Uniform with support $[-y/2\,,(1-y)/2]$ – $\mathcal{U}(-y/2\,,(1-y)/2)$.

All the copula functions selected to carry out the test are one-parameter copulas (δ scalar – $d=1$). At this point, one should note that, in view of the test's robustness to the adopted copula – check (13) – in this selection there was little concern that the copulas would correspond to the above described DGP's for the variate of interest and ME. Rather, and more specifically, the selection was made according to the following criteria: inclusion of the independence copula as a special case for a unique value of the dependence parameter, δ; continuity and differentiability with respect to δ in a neighbourhood of the value yielding independence; analytical convenience, in view of their purported use and the fact that any proper continuous copula can be employed to implement the test – check (13). The copulas' formulae and corresponding expressions for $D(\cdot)$ (defined in (9)) as well as the value of δ yielding independence (denoted δ^i) are detailed next (see, *e.g.*, Joe, 2014, Ch. 4, for these and other examples). In all expressions, as previously defined, $u_Y \equiv F_Y(y)$ and $u_V \equiv F_V(v;\delta)$.

Ali-Mikhail-Haq (AMH)

$$C(u_Y,u_V;\delta)=\frac{u_Y u_V}{1-\delta(1-u_Y)(1-u_V)},\quad |\delta|<1,\quad \delta^i=0;\quad D(u_Y)=2u_Y-1.$$

Farlie-Gumbel-Morgenstern (FGM)

$$C(u_Y,u_V;\delta)=u_Y u_V+\delta u_Y u_V(1-u_Y)(1-u_V),\quad |\delta|<1,\quad \delta^i=0;$$
$$D(u_Y)=2u_Y-1.$$

Frank

$$C(u_Y,u_V;\delta)=-\frac{1}{\delta}\log\left\{1-\frac{[1-\exp(-\delta u_Y)][1-\exp(-\delta u_V)]}{1-\exp(-\delta)}\right\},\quad \delta\in\mathbb{R},\quad \delta^i=0;$$
$$D(u_Y)=u_Y-1/2.$$

Mardia-Takahasi-Clayton-Cook-Johnson (MTCCJ)

$$C(u_Y,u_V;\delta)=\left(u_Y^{-\delta}+u_V^{-\delta}-1\right)^{-1/\delta},\quad \delta\geq 0,\quad \delta^i=0;\quad D(u_Y)=\log u_Y+1.$$

Plackett

$$C(u_Y,u_V;\delta)=\frac{1}{\eta}\left\{1+\eta(u_Y+u_V)-\sqrt{[1+\eta(u_Y+u_V)]^2-4\delta\eta u_Y u_V}\right\},$$
$$\eta\equiv\delta-1,\quad \delta\geq 0,\quad \delta^i=1;\quad D(u_Y)=(2u_Y-1)^3.$$

The expressions of $D(u_Y)$ provide the basis for the test statistic M_n, defined in (12).[(2)] Recalling that the probability integral transform of Y, $U_Y \equiv F_Y(Y)$, is distributed as $\mathcal{U}(0,1)$, one can easily check that all the above expressions for $\mathrm{E}[D(U_Y)]$ have zero expectation – as predicted by (13). The first three functions (AMH, FGM and Frank copulas) lead to the same test statistic, hence, in short, the present Monte Carlo exercise illustrates the performance of three alternative copula-based specification tests of F_Y.

4.2 Measurement Error in a Fractional Variable (no Covariates)

This section reports results illustrating the empirical performance of the proposed test when ME affects a fractional variable (no covariates present). In this part of the experiment, $Y \sim Beta(\alpha;\beta)$ with each of the following values of $(\alpha;\beta)$: Case A – $(3;3)$; B – $(1;5)$; C – $(6;2)$. The conditional distributions used for the fractional variable affected by ME, Z, are as described in the general design.

Table 1.1 illustrates the empirical size of the tests at the 5% nominal size, for cases A through C. The tests based on the MTCCJ copula ($D(u_Y) = \log u_Y + 1$) behave clearly worse than the other two tests, over-rejecting H_0 in all the cases considered. For this reason, the ensuing power estimations do not include MTCCJ copula-based tests. With regard to the other two tests, the M1 test, not surprisingly, outperforms the M2 test in most of the cases. Indeed, with few exceptions, the M2 test displays empirical rejection rates of H_0 that disallow a straightforward use of critical values from the asymptotic χ_1^2 null law. For this reason, only the power estimates for M1 tests are presented.

Table 1.2 displays power estimates for the M1 tests under the adopted DGP's (not including the test based upon the MTCCJ copula). The power results vary considerably across the different DGP's considered in the experiment. One clear hint, though, is that the test based upon the Plackett copula (under which $D(u_Y) = (2u_Y - 1)^3$) seems quite less able than the first test in the table to discern ME in the data. This is particularly the case with a conditional Beta ME distribution, under

(2) The construction of the matrix H, in the estimator $\hat{\Sigma}_n$ of the variance of $\hat{D}_n$, involves evaluation of derivatives of the incomplete beta function at each sample point, z_i. For brevity's sake, the analytic expressions for these derivatives are omitted.

which the amount of ME can be relatively small when compared with the truncated Normal and Uniform error densities. Thus, namely when small amounts of ME are suspected in the data, the test with $D(u_Y) = u_Y - 1/2$ may prove a sound choice.

Table 1.1
Estimated Sizes (%) of M1 and M2 Tests (5% Nominal Size)

	A	B	C
$D = F_Y(Z) - 1/2$			
M1	.052	.054	.054
M2	.058 [a]	.059 [a]	.058 [a]
$D = [2F_Y(Z) - 1]^3$			
M1	.048	.047	.051
M2	.050	.049	.055 [a]
$D = \log F_Y(Z) + 1$			
M1	.061 [a]	.065 [a]	.059 [a]
M2	.064 [a]	.070 [a]	.062 [a]

[a] : 5% rejection rate outside 95% confidence interval.

Table 1.2
Estimated Powers (%) of M1 Test (5% Nominal Size)

	A	B	C
$V + y \vert y \sim Beta(1/(1-y), 2)$			
$D = F_Y(Z) - 1/2$	.300	.067	.813
$D = [2F_Y(Z) - 1]^3$	.051	.047	.068
$V \vert y \sim \mathcal{NT}(0,1)$			
$D = F_Y(Z) - 1/2$	.052	.985	.837
$D = [2F_Y(Z) - 1]^3$	.048	.558	.316
$V \vert y \sim \mathcal{U}(-y/2, (1-y)/2)$			
$D = F_Y(Z) - 1/2$	.047	.995	.868
$D = [2F_Y(Z) - 1]^3$	.047	.612	.403

4.3 Measurement Error in a Fractional Response

This section reports results on the empirical size and power of the proposed test when ME affects a regression's fractional response. The latter is supposed to follow a conditional Beta distribution with parameters involving a single standard normal regressor, $X \sim \mathcal{N}(0,1)$ (newly drawn in each replica). The conditional Beta distribution is parameterized as $Beta(1, 1/\theta)$, with $\theta \equiv G(\beta_1 + \beta_2 x)/[1 - G(\beta_1 + \beta_2 x)]$, so that $\mathrm{E}(Y|X = x) = G(\beta_1 + \beta_2 x)$. Like Y, the function $G(\cdot)$ must be double bounded so it is specified as one of the following alternative models: Logit (Case L), Probit (P), Cauchit (C) and Log-log (LL). These specifications are chosen so as to consider different forms of $G(\cdot)$, namely in terms of symmetry and tail behaviour. As is well known, the Logistic, standard Normal and Cauchy functions are symmetric about $.5$ and, consequently, approach 0 and 1 at the same rate. On the other hand, the loglog model is not symmetric, increasing sharply at small values of $G(\cdot)$ and slowly when $G(\cdot)$ is near 1. The Cauchy distribution presents the heaviest tails, thus being more robust to outliers than the logistic and standard normal formulations. In all the cases considered for $G(\beta_1 + \beta_2 x)$, $(\beta_1; \beta_2)$ are set at $(-1; 0.5)$. The conditional distributions adopted for generating V correspond to those described in the general design.

Table 2.1
Estimated Sizes (%) of M1 and M2 Tests (5% Nominal Size)

	L	P	C	LL
$D = F_Y(Z) - 1/2$				
M1	.052	.054	.056 [a]	.054
M2	.059 [a]	.060 [a]	.061 [a]	.060 [a]
$D = [2F_Y(Z) - 1]^3$				
M1	.053	.053	.051	.051
M2	.058 [a]	.057 [a]	.057 [a]	.057 [a]
$D = \log F_Y(Z) + 1$				
M1	.061 [a]	.056 [a]	.049	.052
M2	.067 [a]	.062 [a]	.066 [a]	.068 [a]

[a] : 5% rejection rate outside 95% confidence interval.

Table 2.1 reports the empirical size of the tests M1 and M2 at the 5% nominal size, for each functional form of the (error-free) conditional mean. Once again, the tests based on the MTCCJ copula (under which $D[F_Y(Z)] = \log F_Y(Z) + 1$) prove unreliable in most cases, with reference to the asymptotic χ_1^2 critical values. A few exceptions to this finding now occur with the M1 tests and Cauchit/Log-log null functional forms, under which either the first test in the table (with $D[F_Y(Z)] = F_Y(Z) - 1/2$) fares worse or the nominal 5% rejection probability is not rejected. Nonetheless, the poor performance of all M2 tests again seems clear, systematically oversized with reference to asymptotic critical values. For this reason, as in Subsection 4.2, only the power estimates of the M1 tests are shown below.

Table 2.2
Estimated Powers (%) of M1 Test (5% Nominal Size)

	L	P	C	LL
$V + y\rfloor y \sim Beta(1/(1-y), 2)$				
$D = F_Y(Z) - 1/2$	.093	.088	.106	.076
$D = [2F_Y(Z) - 1]^3$	.049	.051	.049	.067
$D = \log F_Y(Z) + 1$	.081	.075	.045	.083
$V\rfloor y \sim \mathcal{NT}(0,1)$				
$D = F_Y(Z) - 1/2$	.614	.959	.700	.998
$D = [2F_Y(Z) - 1]^3$	.311	.475	.301	.654
$D = \log F_Y(Z) + 1$	.552	.946	.440	.996
$V\rfloor y \sim \mathcal{U}(-y/2, (1-y)/2)$				
$D = F_Y(Z) - 1/2$	.674	.976	.744	.999
$D = [2F_Y(Z) - 1]^3$	.347	.547	.348	.733
$D = \log F_Y(Z) + 1$	.602	.963	.492	.996

Table 2.2 displays power estimates for the M1 tests. One obvious remark concerns the performance of the tests with predominantly small amounts of ME – the case under the Beta distribution with null conditional mode.[3] In such cases, the low

(3) Similar results were obtained with ME following a conditional truncated normal with null conditional mode.

power of all the variants of the test is noteworthy, namely the Packett copula-based test. On the contrary, power estimates are rather high under more harmful ME, as in the remaining two cases. In this regard, the Plackett copula test again seems to fare worse than the other two (with a cautionary note on the MTCCJ test, given its excessive empirical size with logit and probit functional forms). As in Subsection 4.2 (no covariates), the M1 test based on $D[F_Y(Z)] = F_Y(Z) - 1/2$ appears as the most powerful in every situation considered. The only caveat is now suggested by the case with a null Cauchit conditional mean, under which the test (contrarily to the remaining tests) appears slightly oversized (check Table 2.1).

5 Concluding Remarks

This paper addresses the problem of ME of double bounded random variates, a context which does not allow adoption of classical assumptions concerning ME. Although, for the most part, the text is concerned with fractional variables, the main ideas here suggested can obviously be employed with any continuous double bounded variate of interest. The paper adopts what might be termed an 'indirect' approach to the issue of ME, by examining the dependence between the latter and the unobserved error-free variate of interest. The resulting specification test can be formalized in different ways, depending on the particular bivariate copula which is used to specify these two variables' joint distribution. The simulation results, on the fractional case, suggest that at least one variant of the considered score tests is both reliable and acceptably powerful in the presence of a fractional variable measured with error.

The present text has suggested several hints for future related work. One such hint consists on the adaptation of the present ideas to the development of tests for nonclassical ME of other types of continuous variables (earnings, for example – see, *e.g.*, Gottschalk and Huynh, 2010). Also, the extension of the proposed copula-based procedure to the case of multivariate fractional variables (see, *e.g.*, Mullahy, 2015; Murteira and Ramalho, 2016) appears as a challenging avenue for subsequent research.

José M. R. Murteira.
Address for correspondence: Faculdade de Economia, Universidade de Coimbra, Av. Dias da Silva, 165, 3004-512 Coimbra, Portugal. E-mail: jmurt@fe.uc.pt.

References

Balakrishnan, N., & Lai, C.-D. (2009). *Continuous Bivariate Distributions*. Dordrecht: Springer.

Chesher, A., Dumangane, M., & Smith, R. (2002). Duration Response Measurement Error. *Journal of Econometrics*, *111*, 169-194.
DOI: 10.1016/S0304-4076(02)00103-3.

Chen, X., Hong, H., & Nekipelov, D. (2011). Nonlinear Models of Measurement Errors. *Journal of Economic Literature*, *49*, 901-937.

Ferrari, S., & Cribari-Neto, F. (2004). Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, *31*, 799-815.
DOI: 10.1080/0266476042000214501.

Gottschalk, P. & Huynh, M. (2010). Are Earnings Inequality and Mobility Overstated? The Impact of Nonclassical Measurement Error. *The Review of Economics and Statistics*, *92*, 302-315. DOI: 10.1162/rest.2010.11232.

Hausman, J. (2001). Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *Journal of Economic Perspectives*, *15*, 57-67. DOI: 10.1257/jep.15.4.57.

Joe, H. (2014). *Dependence Modelling with Copulas*. Boca Raton: CRC Press.

Mullahy, J. (2015). Multivariate Fractional Regression Estimation of Econometric Share Models. *Journal of Econometric Methods*, *4*, 71-100.

Murteira, J., & Ramalho, J. (2016). Regression Analysis of Multivariate Fractional Data. *Econometric Reviews*, *35*, 515-552. DOI: 10.1080/07474938.2013.806849.

Newey, W. (1985). Maximum Likelihood Specification Testing and Conditional Moment Tests. *Econometrica*, *53*, 1047-1070. DOI: 10.2307/1911011.

Pagan, A., & Vella, F. (1989). Diagnostic Tests for Models Based on Individual Data: A Survey. *Journal of Applied Econometrics*, *4*, S29-S59.
DOI: 10.1002/jae.3950040504.

Papke, L., & Wooldridge, J. (1996). Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates. *Journal of Applied Econometrics*, *11*, 619-632.

DOI: 10.1002/(SICI)1099-1255(199611)11:6<619::AID-JAE418>3.0.CO;2-1.

Ramalho, E., Ramalho, J., & Murteira, J. (2011). Alternative Estimating and Testing Empirical Strategies for Fractional Regression Models. *Journal of Economic Surveys*, *25*, 19-68. DOI: 10.1111/j.1467-6419.2009.00602.x.

Schennach, S. (2013). Measurement Error in Nonlinear Models – A Review. In D. Acemoglu, M. Arellano & E. Dekel (Eds.), *Advances in Economics and Econometrics*, Econometric Society Monographs 51, 3, Ch. 8. New York: Cambridge University Press.

CeBER Working Papers

(Available on-line at www.uc.pt/go/ceber)

2018-13 *Copula-based Tests for Nonclassical Measurement Error – The Case of Fractional Random Variables - José M. R. Murteira*

2018-12 *Differentiated Impact of Spread Determinants by Personal Loan Category: Evidence from the Brazilian Banking Sector – José Valente, Mário Augusto & José Murteira*

2018-11 *The Effect of Family Ownership, Control and Management on Corporate Debt Structure – Evidence from Panel Fractional Data – Mário Augusto, José Murteira & António Pedro Pinto*

2018-10 *Predictability of stock returns and dividend growth using dividend yields: An international approach – Ana Sofia Monteiro, Hélder Sebastião & Nuno Silva*

2018-09 *Political and institutional determinants of credit booms – Vítor Castro & Rodrigo Martins*

2018-08 *An Aggregate View of Portuguese Exports and Competitiveness – Pedro Bação, António Portugal Duarte & Diogo Viveiros*

2018-07 *The Cycle of recycling and sustainable development. Evidence from the OECD Countries – Pedro André Cerqueira, Elias Soukiazis & Sara Proença*

2018-06 *Information Transmission Between Cryptocurrencies: Does Bitcoin Rule the Cryptocurrency World? – Pedro Bação, António Portugal Duarte, Hélder Sebastião & Srdjan Redzepagic*

2018-05 *Endogenous Growth and Entropy – Tiago Neves Sequeira, Pedro Mazeda Gil & Óscar Afonso*

2018-04 *Determinants of overall and sectoral entrepreneurship: evidence from Portugal – Gonçalo Brás & Elias Soukiazis*

2018-03 *Young and healthy but reluctant to donate blood: An empirical study on attitudes and motivations of university students – Tiago Henriques & Carlota Quintal*

2018-02 *The Iberian electricity market: Price dynamics and risk premium in an illiquid market – Márcio Ferreira & Hélder Sebastião*

2018-01 *Health Investment and Long run Macroeconomic Performance: a quantile regression approach – Francisca Silva, Marta Simões & João Sousa Andrade*

2017-12 *Deflation in the Euro Zone: Overview and Empirical Analysis – Pedro Bação & António Portugal Duarte*

2017-11 *Fiscal Consolidation Programs and Income Inequality – Pedro Brinca, Miguel H. Ferreira, Francesco Franco, Hans A. Holter & Laurence Malafry*

2017-10 *The interconnections between Renewable Energy, Economic Development and Environmental Pollution. A simultaneous equation system approach - Elias Soukiazis, Sara Proença & Pedro André Cerqueira*

2017-09 *The Renminbi: A Warrior for Competitiveness? – Pedro Bação, António Portugal Duarte & Matheus Santos*

2017-08 *Le Portugal et l'Euro – João Sousa Andrade*

2017-07 *The Effect of Public Debt on Growth in Multiple Regimes in the Presence of Long-Memory and Non-Stationary Debt Series - Irina Syssoyeva-Masson & João Sousa Andrade*

2017-06 *The Blank and the Null: An examination of non-conventional voting choices – Rodrigo Martins*

2017-05 *Where is the information on USD/Bitcoins hourly price movements? - Helder Sebastião, António Portugal Duarte & Gabriel Guerreiro*

2017-04 *The response of non-price competitiveness and productivity due to changes in passed income gaps. Evidence from the OECD countries - Pedro André Cerqueira, Micaela Antunes & Elias Soukiazis*

2017-03 *Dutch Disease in Central and Eastern European Countries - João Sousa Andrade & António Portugal Duarte*

2017-02 *On the gains of using high frequency data and higher moments in Portfolio Selection- Rui Pedro Brito, Hélder Sebastião & Pedro Godinho*

2017-01 Growth adjustments through non-price competitiveness and productivity. A cumulative causation approach- Elias Soukiazis, Micaela Antunes & Pedro André Cerqueira

A série CeBER Working Papers foi iniciada em 2017.