Goodness of Link Tests for Multivariate Regression Models

José M. R. Murteira

CEMAPRE, ISEG-Universidade de Lisboa

and

Faculdade de Economia, Universidade de Coimbra.

Address for correspondence:

Av. Dias da Silva, 165, 3004-512 Coimbra, Portugal.

E-mail: jmurt@fe.uc.pt.

September, 2014

Abstract

This note presents an approximation to multivariate regression models which is obtained from a first-order series expansion of the multivariate link function. The proposed approach yields a variable-addition approximation of regression models that enables a multivariate generalization the well-known goodness of link specification test, available for univariate generalized linear models. Application of this general methodology is illustrated with models of multinomial discrete choice and multivariate fractional data, in which context it is shown to lead to well-established approximation and testing procedures.

MSC classification codes: 62J12, 62H15.

Key Words: multivariate regression, multivariate link, variable-addition approximation, goodness of link test.

1 Introduction

This note presents a general variable-addition approximation to multivariate regression models that generalizes the approach introduced by Pregibon (1980), for univariate generalized linear models. The proposed methodology provides a unified framework that encompasses several variable-addition approximation strategies proposed (*e.g.*, in the Econometrics literature).

As described below, the present approximation is obtained from a modification (series expansion) of a particular maintained model of interest, nesting some simpler null model. To the extent that it leads to an added-variables approximation, the approach logically suggests a general specification test that can be implemented by assessing the significance of the additional covariates in the null model. The resulting procedure can thus be implemented without having to estimate the alternative model, which, in the multivariate case, can prove a significant computational advantage on its own.

The remainder of the paper is organized as follows. Section 2 presents the basic variable-addition approximation, developed within a general multivariate regression framework and discusses goodness of link specification tests that are suggested by the proposed approximation. Section 3 applies the proposed approach to several examples from the areas of multinomial discrete choice and multivariate fractional data modelling. Section 4 concludes and suggests subsequent research.

2 Model Approximation and Testing

Let $\boldsymbol{y} \equiv (y_1, \ldots, y_J)'$ denote a random J-vector with conditional expectation, given a matrix of observable covariates $\boldsymbol{X}$, generally expressed as

$$\boldsymbol{G}(\boldsymbol{X};\boldsymbol{\beta}) \equiv [G_1(\boldsymbol{X};\boldsymbol{\beta}), \cdots, G_J(\boldsymbol{X};\boldsymbol{\beta})]' \equiv [E(y_1|\boldsymbol{X}), \cdots, E(y_J|\boldsymbol{X})]' \equiv E(\boldsymbol{y}|\boldsymbol{X}), \quad (1)$$

where $\boldsymbol{\beta}$ denotes a column K-vector of parameters. For simplicity, the conditional means of $\boldsymbol{y}$ may be referred to without explicit mention of their arguments: $\boldsymbol{G} \equiv \boldsymbol{G}(\boldsymbol{X};\boldsymbol{\beta})$ and $G_j \equiv G_j(\boldsymbol{X};\boldsymbol{\beta})$. Frequently, $\boldsymbol{G}$ is an index function, $\boldsymbol{G}(\boldsymbol{X}\boldsymbol{\beta})$, with the forms of $\boldsymbol{G}$, $\boldsymbol{X}$ and $\boldsymbol{\beta}$ depending on the specific context. For instance, in the linear seemingly unrelated regression (SUR) model, typically $G_j(\boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{x}_j'\boldsymbol{\beta}_j$, $j = 1, \ldots, J$, so $\boldsymbol{G}(\cdot)$ is a vector of identity functions, $\boldsymbol{X}$ is block diagonal with each diagonal block given by the row vector

$\boldsymbol{x}'_j$, $j = 1, ..., J$, and $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_J)'$. Another common example arises in the area of discrete choice analysis, where y_j is usually coded as 1 if alternative j is selected (0 otherwise) so $G_j(\boldsymbol{X\beta}) = \Pr(y_j = 1|\boldsymbol{X})$ and $\boldsymbol{G}$ is usually specified as a multivariate c.d.f. (often logistic or normal) – see, *e.g.*, Cameron and Trivedi (2005, Chapter 15).

Model $\boldsymbol{G}$ can often be appropriately nested within a general model, $\boldsymbol{H}(\boldsymbol{X\beta}, \boldsymbol{\eta}) \equiv [H_1(\boldsymbol{X\beta}, \boldsymbol{\eta}), ..., H_J(\boldsymbol{X\beta}, \boldsymbol{\eta})]'$, with $\boldsymbol{\eta}$ an L-vector of parameters such, that $\boldsymbol{H}(\boldsymbol{X\beta}, \boldsymbol{0}) = \boldsymbol{G}(\boldsymbol{X\beta})$. Then, the correctness of $\boldsymbol{G}$ as a functional form for $E(\boldsymbol{y}|\boldsymbol{X})$ can be assessed by testing the hypothesis $H_0 : \boldsymbol{\eta} = \boldsymbol{0}$. Common examples of such nests occur, for instance, in the area of discrete choice with the nested logit and dogit generalizations of the multinomial logit model (respectively, McFadden, 1981, and Gaudry and Dagenais, 1979).

The alternative model can be approximated by including appropriate additional covariates in the null conditional mean function, $\boldsymbol{G}$. One possible approximation stems from consideration of the multivariate extension of the so-called link function, a widely used concept in the generalized linear models (GLM) literature (see Fahrmeir and Tutz, 2001).

The multivariate link is defined as the vector function relating a set of linear predictors to a conformable vector of conditional means. Denote the vector of J conditional means under the alternative hypothesis as

$$\boldsymbol{\mu} = \boldsymbol{H}(\boldsymbol{X\beta}, \boldsymbol{\eta}). \tag{2}$$

If, as is usually the case, the functions $H_j(\boldsymbol{X\beta}, \boldsymbol{\eta})$ are continuously differentiable and injective, one can invert (2) to obtain

$$\boldsymbol{X\beta} = \boldsymbol{h}(\boldsymbol{\mu}, \boldsymbol{\eta}), \tag{3}$$

with $\boldsymbol{h}(\cdot, \cdot)$ denoting the J-variate link function associated with $\boldsymbol{H}$. With $\boldsymbol{\eta} = \boldsymbol{0}$, (2) yields the system of null conditional means; write this as $\boldsymbol{\mu} = \boldsymbol{H}(\boldsymbol{X\beta}, \boldsymbol{0}) = \boldsymbol{G}(\boldsymbol{X\beta})$, with associated link $\boldsymbol{g}(\boldsymbol{\mu}) \equiv \boldsymbol{h}(\boldsymbol{\mu}, \boldsymbol{0})$. Given the invertibility of both systems, there is a one-to-one correspondence between links and functional forms. Thus, as in a univariate setting, the correctness of the multivariate mean specification can be assessed by testing the associated link, $\boldsymbol{h}(\cdot, \cdot)$, in the spirit of Pregibon's (1980) goodness of link test for univariate GLM's.

A first-order Taylor-expansion of $\boldsymbol{h}(\boldsymbol{\mu}, \boldsymbol{\eta})$ around $\boldsymbol{\eta} = \boldsymbol{0}$ yields

$$\boldsymbol{h}(\boldsymbol{\mu}, \boldsymbol{\eta}) \approx \boldsymbol{h}(\boldsymbol{\mu}, \boldsymbol{0}) + \left[\nabla_{\boldsymbol{\eta}'} \boldsymbol{h}(\boldsymbol{\mu}, \boldsymbol{\eta})|_{\boldsymbol{\eta}=\boldsymbol{0}}\right] \times \boldsymbol{\eta} = \boldsymbol{g}(\boldsymbol{\mu}) + \left[\nabla_{\boldsymbol{\eta}'} \boldsymbol{h}(\boldsymbol{\mu}, \boldsymbol{\eta})|_{\boldsymbol{\eta}=\boldsymbol{0}}\right] \times \boldsymbol{\eta}, \tag{4}$$

where $\nabla_{\boldsymbol{\eta}'} \boldsymbol{h}(\boldsymbol{\mu}, \boldsymbol{\eta})|_{\boldsymbol{\eta}=\mathbf{0}}$ denotes evaluation at $\boldsymbol{\eta} = \mathbf{0}$ of the $J \times L$ matrix of derivatives of the J elements of $\boldsymbol{h}(\boldsymbol{\mu}, \boldsymbol{\eta})$ with respect to the row L-vector $\boldsymbol{\eta}'$. The alternative link, $\boldsymbol{h}(\boldsymbol{\mu}, \boldsymbol{\eta})$, can thus be approximated by $\boldsymbol{g}(\boldsymbol{\mu}) = \boldsymbol{X\beta} + \boldsymbol{Z\eta}$, with $\boldsymbol{Z} \equiv - \nabla_{\boldsymbol{\eta}'} \boldsymbol{h}(\boldsymbol{\mu}, \boldsymbol{\eta})|_{\boldsymbol{\eta}=\mathbf{0}}$. Therefore, given the continuity of $\boldsymbol{H}$, the alternative model can be approximated in the neighborhood of the null by $\boldsymbol{H}^*(\boldsymbol{X\beta}, \boldsymbol{\eta}) \equiv \boldsymbol{G}(\boldsymbol{X\beta} + \boldsymbol{Z\eta})$.

Often, system (3) cannot be explicitly produced (notable multivariate exceptions, besides the linear case, being the multinomial logit and dogit models), so one can make use of the Implicit Function Theorem to obtain $\boldsymbol{Z}$. Expressions (2) and (3) yield

$$\boldsymbol{\mu} - \boldsymbol{H}(\boldsymbol{X\beta}, \boldsymbol{\eta}) = \mathbf{0} \Rightarrow -\nabla_{\boldsymbol{\eta}'} \boldsymbol{H}(\boldsymbol{X\beta}, \boldsymbol{\eta}) = \mathbf{0} \Leftrightarrow$$
$$-\underset{J \times J}{\nabla_{(\boldsymbol{X\beta})'} \boldsymbol{H}(\boldsymbol{X\beta}, \boldsymbol{\eta})} \times \underset{J \times L}{\nabla_{\boldsymbol{\eta}'} \boldsymbol{h}(\boldsymbol{\mu}, \boldsymbol{\eta})} - \underset{J \times L}{\nabla_{\boldsymbol{\eta}'} \boldsymbol{H}(\boldsymbol{h}, \boldsymbol{\eta})} = \underset{J \times L}{\mathbf{0}} \Leftrightarrow$$
$$-\nabla_{\boldsymbol{\eta}'} \boldsymbol{h}(\boldsymbol{\mu}, \boldsymbol{\eta}) = \left[\nabla_{(\boldsymbol{X\beta})'} \boldsymbol{H}(\boldsymbol{X\beta}, \boldsymbol{\eta})\right]^{-1} \times \nabla_{\boldsymbol{\eta}'} \boldsymbol{H}(\boldsymbol{X\beta}, \boldsymbol{\eta}). \tag{5}$$

Evaluation under H_0 then yields

$$\boldsymbol{Z} \equiv \left(\nabla_{(\boldsymbol{X\beta})'} \boldsymbol{G}\right)^{-1} \times \nabla_{\boldsymbol{\eta}'} \boldsymbol{H}|_{\boldsymbol{\eta}=\mathbf{0}}. \tag{6}$$

If necessary, $\boldsymbol{Z}$ can be obtained through application of the Cramer rule to (5), which amounts to L linear systems in the columns of $\nabla_{\boldsymbol{\eta}'} \boldsymbol{h}(\boldsymbol{H}, \boldsymbol{\eta})$ (one system per column).

Assume that a random sample of N observations $(\boldsymbol{y}_i, \boldsymbol{X}_i)$, $i = 1, ..., N$, is available to the researcher. The specification test that is suggested by the present framework requires evaluation of $\boldsymbol{Z}$ at first-step $\boldsymbol{\beta}$ estimates, $\hat{\boldsymbol{\beta}}$, so the procedure tests the null hypothesis in the augmented model $\boldsymbol{G}\left(\boldsymbol{X}_i\boldsymbol{\beta} + \hat{\boldsymbol{Z}}_i\boldsymbol{\eta}\right)$, with $\hat{\boldsymbol{Z}}_i \equiv \boldsymbol{Z}\left(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}\right)$. The null model can be estimated through some M-estimation method (*e.g.*, maximum likelihood, ML, quasi-ML, or nonlinear least squares), upon which any of the three classical tests can be used (Wald, likelihood ratio, LR, or score, LM). As $\boldsymbol{Z}$ is absent from the model under H_0, the null distribution of these tests is not affected by the first-step estimation of $\boldsymbol{Z}$. Thus, given present-day software packages, easily handling nonlinear estimation, Wald or LR tests seem fairly easy to use as neither involves estimation of the original alternative specification but only requires estimation of model $\boldsymbol{G}$ twice (respectively, without, and with $\hat{\boldsymbol{Z}}_i$).(1) Even when the values of the parameters under H_0 are on the frontier of the

[1] Even as a regression-based procedure, such as those proposed by Wooldridge (1991), the LM test can be difficult to implement as it requires the construction of many new variables. Therefore, its computational advantage over the other classical tests seems, by now, largely offset.

alternative parameter space (*e.g.*, when the elements of $\boldsymbol{\eta}$ are nonnegative), the estimation of the added variables approximate model does not need to impose this restriction so all three procedures (not only the LM test) retain their usual asymptotic null law. Meanwhile, it is interesting to note, with regard to the LM test, that the same test statistic is obtained, irrespective of whether the original alternative model or its variable addition variant is employed.[2] In other words, the LM test is inherently a variable addition test, with additional covariates appropriately defined by $\boldsymbol{Z}$.

Other variable addition strategies are possible, like the augmentation of the null model with powers and cross-products of the elements of $\boldsymbol{X\beta}$, extending the well-known RESET test (Ramsey, 1969) to nonlinear and/or multivariate models (see Pagan and Vella, 1989, Shukur and Edgerton, 2002, Alkhamisi, *et al.*, 2008). The present approach differs from RESET in that the latter approximates the unspecified $\boldsymbol{H}$ by expanding the inverse function $\boldsymbol{G}^{-1}\left[\boldsymbol{H}(\boldsymbol{X\beta})\right]$ around $\boldsymbol{X\beta} = \boldsymbol{0}$, whereas, here, the null model is augmented with $\boldsymbol{Z}$, not necessarily a polynomial in $\boldsymbol{X\beta}$. In this sense, by choosing appropriate additional covariates, the present test, rather than being used as a general procedure like RESET, can be designed to have power against specific departures from the null.

The present approach can be used in various multivariate contexts. The next Section illustrates its application to a few econometric examples from the areas of discrete choice and fractional data modelling. In these areas, the multinomial logit stands out as a tool of choice, so particular attention is devoted to this model.

3 Illustration: Goodness of Link Tests of Discrete Choice and Fractional Data Models

Data on discrete choice and data on fractional variables share the two basic features of being bounded in the unit interval, $0 \leq y_j \leq 1$, $j = 1, \ldots, J$, and adding up to one, $\sum_{j=1}^{J} y_j = 1$. Understandably, the specifications that are commonly used to model the probability of an individual choosing between mutually exclusive alternatives, $P(y_j|\boldsymbol{X}) = E(y_j|\boldsymbol{X})$, $j = 1, \ldots, J$, may also be employed to describe the conditional means of shares, $E(y_j|\boldsymbol{X})$, in the fractional context (see, *e.g.*, Murteira and Ramalho, 2013, and the references therein).

[2] Note that $\nabla_{\boldsymbol{\eta}'}\boldsymbol{H}^*|_{\boldsymbol{\eta}=\boldsymbol{0}} = \nabla_{(\boldsymbol{X\beta})'}\boldsymbol{G} \times \boldsymbol{Z} = \nabla_{(\boldsymbol{X\beta})'}\boldsymbol{G} \times \left[\nabla_{(\boldsymbol{X\beta})'}\boldsymbol{G}\right]^{-1} \times \nabla_{\boldsymbol{\eta}'}\boldsymbol{H}|_{\boldsymbol{\eta}=\boldsymbol{0}} = \nabla_{\boldsymbol{\eta}'}\boldsymbol{H}|_{\boldsymbol{\eta}=\boldsymbol{0}}$.

Models for both types of data are usually estimated through maximization of a multivariate Bernoulli likelihood, given the specification of the vector of conditional probabilities or means. It should be noted, nonetheless, that maximization of this likelihood yields ML estimates for discrete choice parameters whereas, for fractional data models, it corresponds to QML estimation. As is well known, QML estimation requires robust standard errors of parameters' estimates, a requirement to be taken into account when testing the significance of conditional mean regressors.

3.1 Tests of the Logit

The multinomial logit model can be expressed as $G_j(\boldsymbol{X\beta}) = \exp\left(\boldsymbol{x}_j'\boldsymbol{\beta}\right) / \sum_{l=1}^{J} \exp\left(\boldsymbol{x}_l'\boldsymbol{\beta}\right)$, $j = 1, ..., J$,where $\boldsymbol{X} \equiv (\boldsymbol{x}_1, ..., \boldsymbol{x}_J)'$ is $(J \times K)$, with $\boldsymbol{x}_j$ the column K-vector of covariates associated with alternative j, $j = 1, ..., J$, and $\boldsymbol{\beta}$ a column K-vector of parameters.[3] Any alternative model nesting the logit can be suitably approximated by making use of (6). Let $\boldsymbol{G}^- \equiv (G_1, ..., G_{J-1})'$ and $\boldsymbol{\lambda} \equiv \boldsymbol{X}^-\boldsymbol{\beta}$, where $\boldsymbol{X}^- \equiv (\boldsymbol{x}_1, ..., \boldsymbol{x}_{J-1})'$. The expression of $\boldsymbol{Z}$ involves the inverse of the $(J-1)$-square matrix of derivatives of $\boldsymbol{G}^-$ with respect to $\boldsymbol{\lambda}'$, which can be checked to have typical element $\delta_{jl}G_j^{-1} + G_J^{-1},\ j, l = 1, \ldots, J-1$, where δ_{jl} denotes the Kronecker delta and $G_J = 1 - \sum_{j=1}^{J-1} G_j$. Then, $H_j^*(\boldsymbol{X\beta}, \boldsymbol{\eta}) = G_j(\boldsymbol{X\beta} + \boldsymbol{Z\eta})$, with $\boldsymbol{Z}$ a $(J \times L)$-matrix with j-th row the L-vector $\boldsymbol{z}_j' \equiv G_j^{-1} \left.\boldsymbol{\nabla}_{\boldsymbol{\eta}'} H_{\boldsymbol{j}}\right|_{\boldsymbol{\eta}=\boldsymbol{0}}$, $j = 1, ..., J$.

This result allows the added variables approximation to any alternative model nesting the logit to be generally expressed as

$$H_j^*(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\eta}) = \exp\left(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + G_{ij}^{-1}\boldsymbol{f}_{ij}'\boldsymbol{\eta}\right) \Big/ \sum_{l=1}^{J} \exp\left(\boldsymbol{x}_{il}'\boldsymbol{\beta} + G_{il}^{-1}\boldsymbol{f}_{il}'\boldsymbol{\eta}\right), \tag{7}$$

where, for each sample unit, the vectors $\boldsymbol{f}_{ij}$ satisfy the condition $\sum_{j=1}^{J} \boldsymbol{f}_{ij} = \boldsymbol{0}$. For instance, in the univariate case – $J = 2$ – expression (7) yields

$$H^*(\boldsymbol{x}_i'\boldsymbol{\beta}, \eta) = \exp\left[\boldsymbol{x}_i'\boldsymbol{\beta} + (G_i(1-G_i))^{-1}\boldsymbol{f}_i'\boldsymbol{\eta}\right] \Big/ \left\{1 + \exp\left[\boldsymbol{x}_i'\boldsymbol{\beta} + (G_i(1-G_i))^{-1}\boldsymbol{f}_i'\boldsymbol{\eta}\right]\right\}, \tag{8}$$

for some chosen vector $\boldsymbol{f}_i$. This general expression encompasses various binomial logit generalizations proposed in the literature (see, *e.g.*, Stukel, 1988, and the references therein).

[3] In this form, the model is known as "conditional logit", a term introduced by McFadden (1973). The formulation also allows for constant covariates across alternatives and alternative-specific parameters, by including interactions of alternative-specific indicators with alternative-invariant variables.

Note incidentally that, for chosen L, $\boldsymbol{f}_i' = G_i (1 - G_i) \left[(\boldsymbol{x}_i' \boldsymbol{\beta})^2, ..., (\boldsymbol{x}_i' \boldsymbol{\beta})^{L+1} \right]$ yields an approximation to the alternative model that is used for RESET-type tests.

Some multivariate examples – $J \geq 2$ – are considered in the ensuing text.

Dogit

The dogit model (Gaudry and Dagenais, 1979), can be formally expressed as $H_j \equiv S^{-1} \left(\eta_j + G_j \right)$, $j = 1, ..., J$, with G_j the multinomial logit, $S \equiv 1 + \sum_{j=1}^{J} \eta_j$ and the J parameters η_j are non-negative. The model reduces to the logit under $H_0 : \boldsymbol{\eta} = \mathbf{0}$. For this model, $\boldsymbol{f}_{ij}$ is a J-vector with l-th element $\delta_{lj} - G_{ij}$, $l, j = 1, ..., J$, which yields the approximation[4]

$$H_{ij}^* = \exp \left(\boldsymbol{x}_{ij}' \boldsymbol{\beta} + \eta_j G_{ij}^{-1} \right) / \sum_{l=1}^{J} \exp \left(\boldsymbol{x}_{il}' \boldsymbol{\beta} + \eta_l G_{il}^{-1} \right), \quad j = 1, ..., J. \tag{9}$$

Random Parameters Logit

The random parameters logit, also known as "mixed logit" (McFadden and Train, 2000), generalizes the logit by taking $\boldsymbol{\beta}$ as random (owing to such concerns as, *e.g.*, individual heterogeneity of parameters, measurement errors and/or omission of covariates). Taking expectation with respect to some mixing joint density of $\boldsymbol{\beta}$, the model can be expressed as $H_j = E_{\boldsymbol{\beta}} (G_j)$; following Cox (1983), a Taylor-expansion of H_j around $E(\boldsymbol{\beta})$ yields the small variance approximation $G_j + \frac{1}{2} vech \left(\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}'} G_j \right)' \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ contains the nonredundant elements of $COV(\beta_k, \beta_l)$ and $vech(\cdot)$ denotes the column vector stacking the independent elements of $\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}'} G_j$ evaluated at $E(\boldsymbol{\beta})$. Accordingly, $\boldsymbol{f}_{ij}$ now specializes to $vech \left(\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}'} G_{ij} \right) / 2$, so

$$H_{ij}^* = \frac{\exp \left[\boldsymbol{x}_{ij}' \boldsymbol{\beta} + G_{ij}^{-1} vech \left(\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}'} G_{ij} \right)' \boldsymbol{\eta} \right]}{\sum_{l=1}^{J} \exp \left[\boldsymbol{x}_{il}' \boldsymbol{\beta} + G_{il}^{-1} vech \left(\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}'} G_{il} \right)' \boldsymbol{\eta} \right]}, \quad j = 1, ..., J. \tag{10}$$

For discrete-choice the LM test of $H_0 : \boldsymbol{\eta} = \mathbf{0}$ leads to the general Information Matrix Test statistic (White, 1982), which, as is well-known, can be interpreted as a test for neglected heterogeneity (Chesher, 1984). The fully general random parameters logit naturally encompasses particular models of interest on their own. One special case results by restricting parameters' randomness to the intercepts, in which case one can produce, *e.g.*, a simple "poolability" test in the context of panel data. Cardell (1977), among others, considers the random intercepts multinomial logit for cross-sectional data; Pforr (2011)

[4] For discrete choice and fractional data, LM tests of the logit against the dogit have been proposed by, respectively, Tse (1987) and Murteira, *et al.* (2013).

discusses the panel data version of the model. Let t denote time index and write a logit model for each it-th sample unit as

$$G_{itj} = \exp\left(\boldsymbol{d}_j'\boldsymbol{\alpha} + \boldsymbol{x}_{itj}'\boldsymbol{\beta}\right) / \sum\nolimits_{l=1}^{J} \exp\left(\boldsymbol{d}_l'\boldsymbol{\alpha} + \boldsymbol{x}_{itl}'\boldsymbol{\beta}\right), \quad j = 1, ..., J, \tag{11}$$

where $\boldsymbol{d}_j \equiv (\delta_{j1}, ..., \delta_{jJ-1})'$ is a vector of Kronecker deltas, $\boldsymbol{\alpha} \equiv (\alpha_1, ..., \alpha_{J-1})'$, α_J is set to zero for identification purposes and the vectors $\boldsymbol{x}_{itj}$ do not include constant term. This expression can be seen as nested within a panel data logit specification with individual (time-invariant) effects $\boldsymbol{\alpha}_i \equiv (\alpha_{i1}, ..., \alpha_{i,J-1})'$, with mean $E(\boldsymbol{\alpha}_i) = \boldsymbol{\alpha}$ and some covariance matrix $V(\boldsymbol{\alpha}_i)$. Then, the null hypothesis of interest, $H_0 : V(\alpha_{ij}) = 0, \forall i, j$, implies $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}, \forall i$. Acceptance of H_0 prompts pooled estimation through maximization of the log-likelihood $\log L = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{j=1}^{J} y_{itj} \log G_{itj}$, with T_i the number of time observations for individual i. By restricting $V(\boldsymbol{\alpha}_i)$ to a scalar matrix, $V(\boldsymbol{\alpha}_i) = \eta \boldsymbol{I}_{J-1}$, one can produce a simpler model with similar local behaviour. Then, $L = 1$ and for each it-th unit,

$$\boldsymbol{f}_{itj} = tr\left(\nabla_{\boldsymbol{\alpha}\boldsymbol{\alpha}'} G_{itj}\right)/2 = \frac{1}{2} \times \begin{cases} \gamma_{it} G_{itj} + G_{itj} - 2G_{itj}^2, & j = 1, ..., J-1, \\ \gamma_{it} G_{itJ}, & j = J, \end{cases} \tag{12}$$

with $\gamma_{it} \equiv 2\sum_{l=1}^{J-1} G_{itl}^2 - \sum_{l=1}^{J-1} G_{itl}$. The resulting panel data version of (10) is written as

$$H_{itj}^* = \frac{\exp\left[\alpha_j + \boldsymbol{x}_{itj}'\boldsymbol{\beta} + \eta\left(1 - 2G_{itj}\right)\right]}{\sum_{l=1}^{J-1} \exp\left[\alpha_l + \boldsymbol{x}_{itl}'\boldsymbol{\beta} + \eta\left(1 - 2G_{itl}\right)\right] + \exp\left(\boldsymbol{x}_{itJ}'\boldsymbol{\beta}\right)}, \quad j = 1, ..., J-1, \tag{13}$$

with H_{itJ}^* such, that $\sum_{j=1}^{J} H_{itj}^* = 1$.

Nested Logit

One other logit generalization is provided by the nested logit, the most common member of the "generalized extreme-value" class of models (see, *e.g.*, Train, 2009, Ch. 4, and references therein). The nested logit can be expressed as follows (only two decision levels are considered): suppose that $J > 2$ and the alternatives are distributed into L nonoverlapping nests, N_1, ..., N_L, $L < J$. Suppose that alternative j belongs to nest l; the probability of choosing alternative j – or, in the fractional case, the conditional mean of y_j – can be expressed as $H_j = G_j^l \times S_l$,where

$$G_j^l \equiv \exp\left[\boldsymbol{x}_j'\boldsymbol{\beta}/\left(1 + \eta_l\right)\right] / \sum_{k \in N_l} \exp\left[\boldsymbol{x}_k'\boldsymbol{\beta}/\left(1 + \eta_l\right)\right], \tag{14}$$

$$S_l \equiv \left\{\sum_{k \in N_l} \exp\left[\boldsymbol{x}_k'\boldsymbol{\beta}/\left(1 + \eta_l\right)\right]\right\}^{1+\eta_l} \Bigg/ \sum_{m=1}^{L} \left\{\sum_{k \in N_m} \exp\left[\boldsymbol{x}_k'\boldsymbol{\beta}/\left(1 + \eta_m\right)\right]\right\}^{1+\eta_m}.$$

For this model one obtains $\boldsymbol{f}_{ij}$ in (7) as the L-vector with typical element

$$\nabla_{\eta_l} H_{ij}\big|_{\boldsymbol{\eta}=\mathbf{0}} = G_{ij}\left(\delta_{jl}\left\{\log\left[\sum_{k\in N_l}\exp\left(\boldsymbol{x}'_{ik}\boldsymbol{\beta}\right)\right]-\boldsymbol{x}'_{ij}\boldsymbol{\beta}\right\}-\right. \tag{15}$$
$$\left.\sum_{k\in N_l} G_{ik}\left\{\log\left[\sum_{m\in N_l}\exp\left(\boldsymbol{x}'_{im}\boldsymbol{\beta}\right)\right]-\boldsymbol{x}'_{ik}\boldsymbol{\beta}\right\}\right),\quad l=1,\ldots,L.$$

3.2 Tests of the Dogit

By making use of the previous results for the logit, the present method is easily applied to approximate models that nest the dogit specification.

Denoting the dogit model as D_{ij} and the logit as G_{ij}, one has $\nabla_{\boldsymbol{\lambda}} D_{ij} = S^{-1}\nabla_{\boldsymbol{\lambda}} G_{ij}$ so the added variables approximation to a (dogit nesting) alternative model can be expressed as $H^*_{ij} = S^{-1}\left(\delta_j + G^*_{ij}\right)$, with

$$S \equiv 1+\sum_{j=1}^{J}\delta_j, G^*_{ij} \equiv \tag{16}$$
$$\exp\left(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+G^{-1}_{ij}S\boldsymbol{f}'_{ij}\boldsymbol{\eta}\right)\Big/\sum_{l=1}^{J}\exp\left(\boldsymbol{x}'_{il}\boldsymbol{\beta}+G^{-1}_{il}S\boldsymbol{f}'_{ij}\boldsymbol{\eta}\right)$$

and, as before, $\boldsymbol{f}_{ij} \equiv \nabla_{\boldsymbol{\eta}} H_{ij}\big|_{\boldsymbol{\eta}=\mathbf{0}}$ (obviously, the constant S is irrelevant for the test of H_0). One possible generalization of the dogit would be obtained by considering random δ_j parameters, an alternative easily dealt with by the present approach.

3.3 Tests for Non-nested Models

As suggested by Cox (1961) and further explored by Atkinson (1970), a test of the null specification against a non-nested alternative can be obtained by testing the null against a general model that artificially nests the two competing models. While these authors propose an exponential combination of the two alternative likelihoods, very convenient in the case of Gaussian models, diverse artificial nests seem advantageous if other models are of interest (*e.g.*, Pesaran and Pesaran, 1993, Weeks, 1996, and Santos Silva, 2001).

Here, as in Santos Silva (2001), a linear convex combination of the competing models is used, enabling computation of a simple non-nested specification test. Under this approach, the two competing models for $E\left(\boldsymbol{y}|\boldsymbol{X}\right)$, denoted $\boldsymbol{G}\left(\boldsymbol{X\beta}\right)$ and $\boldsymbol{G}^A\left(\boldsymbol{X\gamma}\right)$, are artificially nested within a general model specified as $H_j\left(\boldsymbol{X};\boldsymbol{\beta},\boldsymbol{\gamma},\eta\right) = \left(1-\eta\right)G_j\left(\boldsymbol{X\beta}\right)+\eta G^A_j\left(\boldsymbol{X\gamma}\right)$, where $\boldsymbol{\gamma}$ denotes regression parameters and η denotes a mixing parameter such, that $0 \leq \eta \leq 1$.[5] Let $\boldsymbol{G}$ denote the null hypothesis, corresponding to $H_0 : \eta = 0$ in the artificial model. Using the results of Section 2, the general model can be approximated in the neighborhood of H_0 by taking the vector of additional covariates, $\boldsymbol{Z} = \left(\nabla_{\boldsymbol{\lambda}'} \boldsymbol{G}^{-}\right)^{-1} \left(\boldsymbol{G}^{-A} - \boldsymbol{G}^{-}\right)$, with $\boldsymbol{G}^{-A}$ defined analogously as $\boldsymbol{G}^{-}$. Thus, as before, the test of H_0 can be cast as a variable addition test within the null specification.

A null multinomial logit, for instance, can be tested against a non-nested alternative by considering

$$H_{ij}^{*} = \exp\left[\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \eta\left(G_{ij}^{A}/G_{ij} - 1\right)\right] \Big/ \sum\nolimits_{l=1}^{J} \exp\left[\boldsymbol{x}_{il}'\boldsymbol{\beta} + \eta\left(G_{il}^{A}/G_{il} - 1\right)\right]. \qquad (17)$$

In the univariate case, a test of the logit against, *e.g.*, the probit, $\Phi_i \equiv \Phi\left(\boldsymbol{x}_i'\boldsymbol{\gamma}\right)$ ($\Phi(\cdot)$: standard normal c.d.f.), tests omission of the covariate $\left[G_i\left(1 - G_i\right)\right]^{-1}\left(\Phi_i - G_i\right)$ in the augmented logit. Reversing roles, the test of the probit against the logit (now G_i^A) evaluates the η estimate in $H_i^* = \Phi\left[\boldsymbol{x}_i'\boldsymbol{\beta} + \eta\phi_i^{-1}\left(G_i^A - \Phi_i\right)\right]$, with $\phi_i \equiv \phi\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)$ the standard normal density.

As another example, consider a test of the Dogit against some non-nested alternative (*e.g.*, probit or nested logit). From Section 3.2, the approximation to the artificial model can be written as $H_{ij}^* = S^{-1}\left(\delta_j + G_{ij}^*\right)$, where

$$G_{ij}^{*} \equiv \exp\left[\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \eta\left(G_{ij}^{A} - D_{ij}\right)G_{ij}^{-1}\right] \Big/ \sum_{l=1}^{J} \exp\left[\boldsymbol{x}_{il}'\boldsymbol{\beta} + \eta\left(G_{il}^{A} - D_{il}\right)G_{il}^{-1}\right]. \qquad (18)$$

4 Conclusion

This note presents an approximation to multivariate regression models that is inspired by the goodness of link approach introduced by Pregibon (1980), for univariate GLM's. The suggested approach provides a comprehensive framework which encompasses, as special cases, several approximation and testing procedures well established in the literature. In a multivariate context, this basic idea is particularly attractive as it leads to specification tests that are easily implemented through Wald or LR procedures that only require estimation of the null model, augmented with specific covariates.

The present text suggests some ideas for future research. One such route generalizes the proposed method by extending the Taylor-expansion of the alternative link function to

[5] With appropriate exclusion restrictions on the parameters of both models, different sets of regressors can be included in each conditional expectation.

higher-than-one degree polynomials. This strategy will lead to extended variable-addition specification tests, presumably more powerful against specific alternatives of interest.

Acknowledgements

I thank Joaquim Ramalho and João Santos Silva for helpful comments – remaining errors are obviously covered by the usual disclaimer. Financial support from Fundação para a Ciência e Tecnologia (grant PEst-OE/EGE/UI0491/2013) is gratefully acknowledged.

References

Alkhamisi, M., Khalaf G., Shukur, G. (2008), "The Effect of Fat-tailed Error Terms on the Properties of System-wise RESET Test", *Journal of Applied Statistics*, 35(1), 101-113.

Atkinson, A. C. (1970), "A Method for Discriminating Between Models", *Journal of the Royal Statistical Society Series B*, 32, 323-35.

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

Cardell, S. (1977), "Multinomial Logit Models with Correlated Stochastic Terms", Working paper, Charles River Associates, Boston, MA.

Chesher, A. (1984), "Testing for Neglected heterogeneity", *Econometrica*, 52(4), 865-872.

Cox, D. R. (1961), "Tests of Separate Families of Hypotheses", *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 105–123.

Cox D. R. (1983), "Some Remarks on Overdispersion", *Biometrika*, 70, 269-274.

Fahrmeir, L., Tutz, G. (2001), *Multivariate Statistical Modelling Based on Generalized Linear Models, 2nd ed.*, Springer, New York.

Gaudry, M. J., Dagenais M. G. (1979), "The Dogit Model", *Transportation Research*, 13B(2), 105-111.

McFadden D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior", in: P. Zarembka (Ed.), *Frontiers in Econometrics*, Academic Press, New York.

McFadden D. (1981), "Econometric Models of Probabilistic Choice", in: Manski C., McFadden, D. (Eds.), *Structural Analysis of Discrete Data and Econometric Applications*, MIT Press, Cambridge, MA. pp. 198-272.

McFadden D., Train, K. (2000), "Mixed MNL Models for Discrete Response", *Journal of Applied Econometrics*, 15, 447-470.

Murteira, J., Ramalho, J. (2013), "Regression Analysis of Multivariate Fractional Data", *Econometric Reviews*, forthcoming.

Pagan, A., Vella, F. (1989), "Diagnostic Tests for Models Based on Individual Data: A Survey", *Journal of Applied Econometrics*, 4, 29-59.

Pesaran M. H., Pesaran, B. (1993), "A Simulation Approach to the Problem of Computing Cox's Statistic for Testing Non-nested Models", *Journal of Econometrics*, 57, 377-392.

Pforr, K. (2011), "Implementation of a Multinomial Logit Model with Fixed Effects", presented at: Ninth German Stata Users Group Meeting, Bamberg.

Pregibon, D. (1980), "Goodness of Link Tests for Generalized Linear Models", *Applied Statistics*, 29(1), 15–24.

Ramsey, J. B. (1969), "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis", *Journal of the Royal Statistical Society Series B*, 31, 350-371.

Santos Silva, J. M. C. (2001), "A Score Test for Non-nested Hypotheses with Applications to Discrete Data Models", *Journal of Applied Econometrics*, 16, 577-597.

Shukur, G., Edgerton, D. (2002), "The Small Sample Properties of the Reset Test as Applied to Systems of Equations", *Journal of Statistical Computation and Simulation*, 72(12), 909-924.

Stukel, T. (1988), "Generalized Logistic Models", *Journal of the American Statistical Association*, 83(402), 426-431.

Train, K. E. (2009), *Discrete Choice Methods with Simulation, 2nd. ed.*, Cambridge University Press, Cambridge, UK.

Tse, Y. K. (1987), "A Diagnostic Test for the Multinomial Logit Model", *Journal of Business and Economics Statistics*, 16, 283-286.

Weeks, M. (1996), "Testing the Binomial and Multinomial Choice Models Using Cox's Non-Nested Test", *Journal of the American Statistical Association (Papers and Proceedings),* 105, 519-530.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, 50(1), 1-25.

Wooldridge, J. M. (1991), "On the Application of Robust, Regression-based Diagnostics to Models of Conditional Means and Conditional Variances", *Journal of Econometrics*, 47, 5-46.