Economics Letters 186 (2020) 108769

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Semiparametric quasi maximum likelihood estimation of the fractional response model

Santiago Montoya-Blandón, David T. Jacho-Chávez *

Department of Economics, Emory University, Rich Building, 1602 Fishburne Dr., Atlanta, GA 30322-2240, USA

ARTICLE INFO

Article history:
Received 11 June 2019
Received in revised form 26 September 2019
Accepted 16 October 2019
Available online 21 October 2019

JEL classification:
C21
C25

Keywords:
Fractional response
Semiparametric estimation
Quasi maximum likelihood
Misspecification

ABSTRACT

This paper proposes a new semiparametric estimator of models where the response random variable is a fraction. The estimator is constructed by optimizing a semiparametric quasi-maximum likelihood that utilizes kernel smoothing. Under suitable conditions, the consistency and asymptotic normality of the proposed estimator is established allowing for data-driven bandwidth choices as well as random trimming, and its flexibility and robustness are showcased in a Monte Carlo experiment and an empirical application.

1. Introduction

Fractional response models arise naturally in the study of compositional data that concern a wide array of fields, such as biology, chemistry, economics, geology, and many others (Aitchison, 2003; Kieschnick and McCullough, 2003). The defining characteristic of these data is that they lie within the d-dimensional simplex $\mathcal{S}^d = \left\{(y^{(1)}, \ldots, y^{(d)}) \in \mathbb{R}^d : 0 \leq y^{(j)} \leq 1, j = 1, \ldots, d; \sum_{j=1}^{d} y^{(j)} = 1\right\}$. When one is interested in their modeling as dependent variables, several estimation methods have been proposed in the statistic and econometrics literature for dealing with this data structure in a variety of settings. Some of these frameworks include parametric likelihood methods like the logistic-normal regression (Aitchison and Shen, 1980; Allenby and Lenk, 1994), beta regression (Ferrari and Cribari-Neto, 2004), Dirichlet regression (Mullahy, 2015; Murteira and Ramalho, 2016); quasi-likelihood methods (Papke and Wooldridge, 1996, 2008); two-part models (Cook et al., 2008; Ramalho and Silva, 2009; Stavrunova and Yerokhin, 2012), among others (Ramalho and Ramalho, 2017). Many applications lend themselves to the use of this approach, such as demand estimation (Woodland, 1979; Koch, 2015; Velásquez-Giraldo et al., 2018), firm analysis (Loudermilk, 2007; Sosa, 2009) and finance (Ramalho and Silva, 2009; Stavrunova and Yerokhin, 2012).

In the case of $d = 2$, where only one fraction needs to be modeled, we propose a kernel-based semiparametric quasi-maximum likelihood estimator (SPQMLE) which adapts Papke and Wooldridge's (1996) estimator to an unknown link function. The proposed adaptation inherits the nice properties of the original estimator, such as dealing with boundary values—where the response variable is allowed to take values exactly equal to 1 or 0—and it is robust to potential misspecification in the link function. Furthermore, the asymptotic properties are derived allowing for data-dependent smoothing parameters as well as possible random trimming. By deriving the exact formula of the asymptotic variance–covariance matrix for the proposed SPQMLE it is shown that there is no estimation effect from replacing the unknown link function by a consistent nonparametric kernel estimator.

A Monte Carlo experiment provides evidence that our method performs well in small-sample settings, and this performance is comparable to the performance achieved by a benchmark maximum likelihood estimation method (MLE) and a correctly specified quasi-likelihood method, but uniformly dominates methods with a misspecified link function. An empirical implementation of the proposed estimator utilizing data from Papke and Wooldridge (1996) is also included. Our point estimates are numerically smaller than those originally obtained in Papke and Wooldridge (1996) and closer to the baseline linear regression model.

The remainder of the paper is organized as follows: Section 2 introduces the estimator along with its asymptotic properties,

* Corresponding author.
E-mail addresses: smonto2@emory.edu (S. Montoya-Blandón), djachocha@emory.edu (D.T. Jacho-Chávez).

https://doi.org/10.1016/j.econlet.2019.108769

Section 3 presents the results of our Monte Carlo simulation comparing our method with other suitable candidates, while Section 4 presents the results of our empirical application, and Section 5 concludes.

2. Estimator and asymptotic properties

2.1. Estimator

Assume one has access to an independent and identically distributed (i.i.d.) sample $\{\boldsymbol{y}_i', \boldsymbol{x}_i'\}_{i=1}^n$ from the joint distribution of $(\boldsymbol{Y}', \boldsymbol{X}')$ where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are k and d dimensional random vectors respectively. We will assume that $\boldsymbol{Y}$ takes values in $\mathcal{S}^2$. Note that in this case, one can focus the modeling strategy on one of the components of $\boldsymbol{Y}$ as the other will then be fully determined. Specifically, we will center our attention on $\boldsymbol{Y}^{(1)}$, which we will hereafter denote simply as Y. Given the characteristics of the data discussed before, we introduce the SPQMLE framework. Let the following index restriction holds almost surely (a.s.)

$$\mathrm{E}[Y_i|\boldsymbol{x}_i] = \mathrm{E}[Y_i|\boldsymbol{x}_i'\boldsymbol{\beta}_0] \equiv m(\boldsymbol{x}_i'\boldsymbol{\beta}_0) \tag{1}$$

for some $\boldsymbol{\beta}_0 \in \mathcal{B} \subset \mathbb{R}^p$ and $\boldsymbol{x}_i \in \mathcal{X} \subset \mathbb{R}^p$, where $\mathcal{X}$ represents the support of $\boldsymbol{X}$. We assume $f(\boldsymbol{x}|z)$ is the density of $\boldsymbol{X}$ conditional on $z = \boldsymbol{X}'\boldsymbol{\beta}$ with respect to a measure μ. Our estimator for $\boldsymbol{\beta}_0$ is based on the semiparametric quasi-likelihood function

$$\mathcal{L}_n(\boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{i=1}^{n}\{y_i \log[\widehat{m}(\boldsymbol{x}_i'\boldsymbol{\beta})] + (1-y_i)\log[1-\widehat{m}(\boldsymbol{x}_i'\boldsymbol{\beta})]\}\widehat{t}_{ni}\,, \tag{2}$$

where $\widehat{m}(\boldsymbol{x}_i'\boldsymbol{\beta})$ estimates the conditional mean $M(\boldsymbol{x}_i'\boldsymbol{\beta}) = \mathrm{E}[m(\boldsymbol{x}_i'\boldsymbol{\beta}_0)|\boldsymbol{x}_i'\boldsymbol{\beta}]$, using a (leave-one-out) Nadaraya–Watson estimator given by $\widehat{m}(\boldsymbol{x}_i'\boldsymbol{\beta}) = \widehat{G}(\boldsymbol{x}_i'\boldsymbol{\beta})/\widehat{f}(\boldsymbol{x}_i'\boldsymbol{\beta})$, where $\widehat{G}(\boldsymbol{x}_i'\boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{j\neq i}^n y_j K_{\widehat{h}_n}(\boldsymbol{x}_j'\boldsymbol{\beta} - \boldsymbol{x}_i'\boldsymbol{\beta})$, $\widehat{f}(\boldsymbol{x}_i'\boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{j\neq i}^n K_{\widehat{h}_n}(\boldsymbol{x}_j'\boldsymbol{\beta} - \boldsymbol{x}_i'\boldsymbol{\beta})$ with $K_h(v) = h^{-1}K(v/h)$, $K(\cdot)$ a kernel function, and $\widehat{h}_n$ a possibly data-dependent bandwidth. As the dependent variable in this setting is not binary but a fraction, the likelihood defined in (2) is inherently misspecified (even with a correctly specified fixed $m(\cdot)$ function), and thus consistent estimation is guaranteed by the index restriction in (1) and the conditions given in Theorem 1 (see Papke and Wooldridge, 1996, for possible optimality properties of this quasi-likelihood in the class of the linear exponential family). Let $\mathbb{I}\{\cdot\}$ be the indicator function that equals 1 when its argument is true, and 0 otherwise. Then, $\widehat{t}_{ni} \equiv \mathbb{I}\{\widehat{f}(\boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}}) \geq \tau_n\}$ is a trimming function based on a preliminary consistent estimator of $\boldsymbol{\beta}_0$, denoted by $\widetilde{\boldsymbol{\beta}}$, and $\tau_n \to 0$ as $n \to \infty$ at a rate satisfying Assumption 8 below. This estimator could be obtained, for example, by maximizing (2) using $\widehat{t}_{ni} = \mathbb{I}\{\boldsymbol{x}_i \in A\}$, where $A \in \mathcal{X}$ is a compact subset. The proposed estimator is then given by

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}\in\mathcal{B}} \mathcal{L}_n(\boldsymbol{\beta})\,. \tag{3}$$

2.2. Asymptotic properties

We apply the results in Gourieroux et al. (1984) and Escanciano et al. (2014) to show that our estimator of $\boldsymbol{\beta}_0$ in (1) defined by (2)–(3) is consistent and asymptotically normal. We begin by listing the required assumptions, which set up the model and are needed to guarantee the properties of kernel estimated functions. Throughout, C will denote a generic positive constant that is not necessarily the same.

Assumption 1 (*Identification of $\boldsymbol{\beta}_0$*). (i) there are no constant elements in $\boldsymbol{x}$, (ii) the first element of $\boldsymbol{x}$, say x_1 is continuous and its associated component of $\boldsymbol{\beta}_0$, say $\beta_1 = 1$, and (iii) if $m(\boldsymbol{x}'\boldsymbol{\beta}_1) = m(\boldsymbol{x}'\boldsymbol{\beta}_2)$ a.s. (with respect to the measure μ) then $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ (these are standard in single index models, see for example Ichimura, 1993; Klein and Spady, 1993 and Li and Racine, 2007, pp. 251–253).

The following four assumptions are standard and limit the general set up (Assumptions 2–3), introduce a general rth-order kernel (Assumption 4) and control the bias present in the nonparametric estimations (Assumption 5).

Assumption 2. The observations $\{y_i, \boldsymbol{x}_i'\}_{i=1}^n$ are an i.i.d. sample from the joint distribution of $(Y, \boldsymbol{X}')$, satisfying $\mathrm{E}[|Y|^{2+\delta}|\boldsymbol{X} = \boldsymbol{x}] < \infty$ for almost all $x \in \mathcal{X}$ and some $\delta > 0$.

Assumption 3. $\mathcal{B}$ is a compact set, and $\boldsymbol{\beta}_0 \in \mathrm{int}(\mathcal{B})$.

Assumption 4. The kernel function $K : \mathbb{R} \to \mathbb{R}$ is bounded, symmetric, twice continuously differentiable and satisfies: $\int K(v)dv = 1$, $\int v^l K(v)dv = 0$ for $0 < l < r$, and $\int |v^r K(v)|dv < \infty$ for some $r \geq 2$. Letting $d^{(j)}K(v)/dv^j$ denote the jth derivative of $K(\cdot)$, we further assume that for $j = 1, 2$, $|d^{(j)}K(v)/dv^j| \leq C$, and for some $s > 1$, $|d^{(j)}K(v)/dv^j| \leq C|v|^{-s}$ for $|v| > L_j$, $0 < L_j < \infty$.

Assumption 5. For all $\boldsymbol{\beta}$ and $\boldsymbol{x} \in \mathcal{X}$, $f(\boldsymbol{x}'\boldsymbol{\beta})$, $m(\boldsymbol{x}'\boldsymbol{\beta})$, and $f(\boldsymbol{x}|z)$ are r-times continuously differentiable in $z = \boldsymbol{x}'\boldsymbol{\beta}$, with all functions and derivatives being uniformly bounded.

Assumption 6. The possibly data-dependent bandwidth $\widehat{h}_n$ satisfies $P_n(a_n \leq \widehat{h}_n \leq b_n) \to 1$ as $n \to \infty$, for deterministic sequences of positive numbers a_n and b_n such that $b_n \to 0$, $b_n^{2r} n \to 0$ and $a_n^3 n/\log n \to \infty$, for r as given by Assumption 4.

The final assumptions adapt those in Escanciano et al. (2014) (specifically, see their Assumptions 5, B.7, B.8, and C.1) to guarantee uniform convergence of the estimated functions and their derivatives while allowing for data-dependent bandwidths such as those obtained by plug-in rules and cross-validation (Andrews, 1995), as well as deal with random trimming. Let $\widehat{t}_{ni} \equiv \mathbb{I}\{\boldsymbol{x}_i \in \widehat{\mathcal{X}}_n\}$ represent a trimming function where $\widehat{\mathcal{X}}_n \subset \mathcal{X}$ could potentially be the result of an estimation procedure, such as a subset based on values of $\widehat{f}$. Let $\mathcal{X}_n$ represent a deterministic set and define $t_{ni} \equiv \mathbb{I}\{\boldsymbol{x}_i \in \mathcal{X}_n\}$, as well as the rate $d_n \equiv (\max\{\log 1/a_n, \log\log n\}/a_n n)^{1/2} + b_n^r$.

Assumption 7. The following two conditions are satisfied: (i) there is a sequence τ_n of positive numbers satisfying $\tau_n \leq \inf_{\boldsymbol{\beta}\in\mathcal{B},\boldsymbol{x}\in\mathcal{X}_n} f(\boldsymbol{x}'\boldsymbol{\beta})$, $d_n^4 n/\tau_n^6 \to 0$ and $d_n/\tau_n \to 0$; and (ii) $P_n(\boldsymbol{X}_i \in \mathcal{X}_n) \to 1$ as $n \to \infty$ and $\mathrm{E}[|\widehat{t}_{ni} - t_{ni}|] = o(n^{-1/2})$.

Finally, in order to ensure that the estimated conditional mean asymptotically belongs to a sufficiently well-behaved class, we can further introduce the rate $d_{mn} \equiv (\max\{\log 1/a_n, \log\log n\}/a_n^3 n)^{1/2}$.

Assumption 8. The rate d_{mn} is such that $d_{mn} = O(1)$.

The main result of the paper is summarized by the following theorem (a corresponding outline for the proof can be found in the supplemental material)

Theorem 1. *Given Assumptions 1–5 and 6–8, $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1})$, where*

$$\boldsymbol{A} = E\left\{\frac{m'(\boldsymbol{X}'\boldsymbol{\beta}_0)^2}{m(\boldsymbol{X}'\boldsymbol{\beta}_0)[1-m(\boldsymbol{X}'\boldsymbol{\beta}_0)]}(\boldsymbol{X} - E[\boldsymbol{X}|\boldsymbol{X}'\boldsymbol{\beta}_0]) \times(\boldsymbol{X} - E[\boldsymbol{X}|\boldsymbol{X}'\boldsymbol{\beta}_0])'\right\}, \tag{4}$$