Computational Statistics and Data Analysis 108 (2017) 97–120

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Nonparametric estimation of dynamic discrete choice models for time series data

Byeong U. Park[a], Léopold Simar[b], Valentin Zelenyuk[c,*]

[a] Department of Statistics, Seoul National University, Republic of Korea

[b] Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Belgium

[c] School of Economics and Centre for Efficiency and Productivity Analysis (CEPA) at The University of Queensland, Australia

ARTICLE INFO

Article history:
Received 21 May 2015
Received in revised form 25 October 2016
Accepted 25 October 2016
Available online 11 November 2016

Keywords:
Nonparametric quasi-likelihood
Local-likelihood
Dynamic probit
Forecasting

ABSTRACT

The non-parametric quasi-likelihood method is generalized to the context of discrete choice models for time series data, where the dynamic aspect is modeled via lags of the discrete dependent variable appearing among regressors. Consistency and asymptotic normality of the estimator for such models in the general case is derived under the assumption of stationarity with strong mixing condition. Monte Carlo examples are used to illustrate performance of the proposed estimator relative to the fully parametric approach. Possible applications for the proposed estimator may include modeling and forecasting of probabilities of whether a subject would get a positive response to a treatment, whether in the next period an economy would enter a recession, or whether a stock market will go down or up, etc.

1. Introduction

Discrete response (or choice) models have received substantial interest in many areas of research. Since the influential works of McFadden (1973, 1974) and Manski (1975), these models have become very popular in economics, especially microeconomics, where they were elaborated on and generalized in many respects. Some very interesting applications of such models are also found in macroeconomic studies where one needs to take into account time series aspects of data. Typical applications of the time series discrete response models deal with forecasting of economic recessions, the decisions of central banks on interest rate, movements of the stock market indices, etc. (See Estrella and Mishkin, 1995, 1998, Dueker, 1997, 2005, Russell and Engle, 1998, 2005, Park and Phillips, 2000, Hu and Phillips, 2004, Chauvet and Potter, 2005, Kauppi and Saikkonen, 2008, de Jong and Woutersen, 2011, Harding and Pagan, 2011, Kauppi, 2012 and Moysiadis and Fokianos, 2014 to mention just a few.)

The primary goal of this work is to develop a methodology for non-parametric estimation of dynamic time series discrete response models, where the discrete dependent variable is related to its own lagged values as well as other regressors. The theory we develop in the next two sections is fairly general and can be used in many areas of research.

The reason for going non-parametric, at least as a complementary approach, is very simple, yet profound: The parametric maximum likelihood in general, and probit or logit approaches in particular, yield inconsistent estimates if the parametric assumptions are misspecified. Many important works addressed this issue in different ways, e.g., see Cosslett (1983, 1987),

* Correspondence to: 530, Colin Clark Building (39), St Lucia, Brisbane, Qld 4072, Australia.
E-mail address: v.zelenyuk@uq.edu.au (V. Zelenyuk).

http://dx.doi.org/10.1016/j.csda.2016.10.024
0167-9473/

Manski (1985), Klein and Spady (1993), Horowitz (1992), Matzkin (1992, 1993), Fan et al. (1995), Lewbel (2000), Honore and Lewbel (2002), Frölich (2006), Dong and Lewbel (2011) and Harding and Pagan (2011), to mention just a few.

The main contribution of our work to existing literature is that we generalize the method of Fan et al. (1995) to the context that embraces time series aspects and in particular the case with lags of the (discrete) dependent variable appearing among the regressors. Such a dynamic feature of the model is very important in practice. For example, in weather forecasting, one would also naturally expect that the lagged dependent variable, describing whether the previous day was rainy or not, may play a very important role in explaining the probability that the next day will also be rainy. Another example of the importance of the dynamic component among the explanatory factors in discrete response models can be found in the area of forecasting economic recessions (Dueker, 1997 and Kauppi and Saikkonen, 2008).

We derive the asymptotic theory for the proposed estimator under the assumption of stationarity with strong mixing condition (in the spirit of Masry, 1996). Our approach is different from and compliments to another powerful non-parametric approach based on the Nadaraya–Watson estimator (e.g., see Harding and Pagan, 2011). Specifically, we use an alternative estimation paradigm – the one based on the non-parametric quasi-likelihood and the local likelihood concepts – which have well-known advantages over the least squares approach for the context of discrete response models. Furthermore, we consider and derive the theory for the local linear fit, which is known to provide more accurate estimation of a model than the local constant approach and is more convenient for estimation of derivatives or marginal effects of the regressors on (the expected value of) the response variable.

It is also worth noting here that a related approach (a special case of ours) was used by Frölich (2006) who considered the local likelihood method in the case of a binary logit-type regression model with both continuous and discrete explanatory variables, in a cross-section set up. Specifically, Frölich (2006) provided very useful and convincing Monte Carlo evidence about superior performance of the local likelihood logit relative to parametric logit for his set up (cross-sectional), but without deriving asymptotic properties of the resulting estimators. Our work encompasses the work of Frölich (2006) as a special case, and, importantly, allows for time series nature of the data, including the dynamic aspect, and provides key asymptotic results for this set up that appears to be missing in the literature. A natural extension to our work would be to also allow for non-stationarity (e.g., as in Park and Phillips, 2000), which is a subject in itself and so we leave it for future research.

Our paper is structured as following: Section 2 outlines the general methodology, Section 3 outlines the theoretical properties of the proposed estimator, Section 4 discusses the choice of bandwidths, Section 5 provides some Monte Carlo evidence, while the Appendix provides further details.

2. General methodology

Suppose we observe $(\mathbf{X}^i, \mathbf{Z}^i, Y^i),\ 1 \le i \le n$, where $\{(\mathbf{X}^i, \mathbf{Z}^i, Y^i)\}_{i=-\infty}^{\infty}$ is a stationary random process. We assume that the process satisfies a strong mixing condition, as described in detail in the next section. The response variable Y^i is of discrete type. For example, it may be binary taking the values 0 and 1. The vector of covariates $\mathbf{X}^i$ is of d-dimension and of continuous type, while $\mathbf{Z}^i$ is of k-dimension and of discrete type. The components of the vector Z^i are allowed to be lagged values of the response variable. For example, $\mathbf{Z}^i = (Y^{i-1}, \ldots, Y^{i-k})$. Our main interest is to estimate the mean function

$$m(\mathbf{x}, \mathbf{z}) = E(Y^i | \mathbf{X}^i = \mathbf{x}, \mathbf{Z}^i = \mathbf{z}).$$

We employ the quasi-likelihood approach of Fan et al. (1995) to estimate the mean function. It requires two ingredients. One ingredient is the specification of a quasi-likelihood $Q(\cdot, y)$, which is understood to take the role of the likelihood of the mean when $Y = y$ is observed. It is defined by $\partial Q(\mu, y)/\partial \mu = (y - \mu)/V(\mu)$, where V is a chosen function for the working conditional variance model $\sigma^2(\mathbf{x}, \mathbf{z}) \equiv \mathrm{var}(Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) = V(m(\mathbf{x}, \mathbf{z}))$, where here and below $(\mathbf{X}, \mathbf{Z}, Y)$ denotes the triple that has the same distribution as $(\mathbf{X}^i, \mathbf{Z}^i, Y^i)$. The other ingredient is the specification of a link function g. The link function should be strictly increasing. In a parametric model where it is assumed that $g(m(\mathbf{x}, \mathbf{z}))$ takes a parametric form, its choice is a part of the parametric assumptions. Thus, a wrong choice would jeopardize the estimation of m. In nonparametric settings, its choice is less important. One may take simply the identity function as a link, but one often needs to use a different one. One case is where the target function m has a restricted range, such as the one where Y is binary so that m has the range $[0, 1]$. A proper use of a link function guarantees the correct range.

With a link function g and based on the observations $\{(\mathbf{X}^i, \mathbf{Z}^i, Y^i)\}_{i=1}^{n}$, the quasi-likelihood of the function f defined by $f(\mathbf{x}, \mathbf{z}) = g(m(\mathbf{x}, \mathbf{z}))$ is given by $\sum_{i=1}^{n} Q(g^{-1}(f(\mathbf{X}^i, \mathbf{Z}^i)), Y^i)$. Let $(\mathbf{x}, \mathbf{z})$ be a fixed point of interest at which we want to estimate the value of the mean function m or the transformed function f. We apply a local smoothing technique to the observations $(\mathbf{X}^i, \mathbf{Z}^i)$ near $(\mathbf{x}, \mathbf{z})$. In the space of the continuous covariates the weights applied to the data points change smoothly on the scale of the distance to the point $(\mathbf{x}, \mathbf{z})$, while in the space of discrete covariates they take some discrete values, one for the case $\mathbf{Z}^i = \mathbf{z}$ and the others for $\mathbf{Z}^i \ne \mathbf{z}$. Specifically, we use a product kernel $w_c^i(\mathbf{x}) \times w_d^i(\mathbf{z})$ for the weights of $(\mathbf{X}^i, \mathbf{Z}^i)$ around $(\mathbf{x}, \mathbf{z})$, where

$$w_c^i(\mathbf{x}) = \prod_{j=1}^{d} K_{h_j}(x_j, X_j^i), \qquad w_d^i(\mathbf{z}) = \prod_{j=1}^{k} \lambda_j^{I(Z_j^i \ne z_j)}.$$