Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems 157 (2006) 2579 – 2592
FUZZY sets and systems
www.elsevier.com/locate/fss

Least-squares estimation in linear regression models with vague concepts

Volker Krätschmer*

Statistics and Econometrics, Faculty of Law and Economics, University of Saarland, D-66041 Saarbrücken, Germany

Available online 5 June 2006

Abstract

The paper is a contribution to parameter estimation in fuzzy regression models with random fuzzy sets. Here models with crisp parameters and fuzzy observations of the variables are investigated. This type of regression models may be understood as an extension of the ordinary single equation linear regression models by integrating additionally the physical vagueness of the involved items. So the significance of these regression models is to improve the empirical meaningfulness of the relationship between the items by a more sensitive attention to the fundamental adequacy problem of measurement. Concerning the parameter estimation the ordinary least-squares method is extended. The existence of estimators by the suggested method is shown, and some of their stochastic properties are surveyed.

Keywords: Problem of adequacy; Physical vagueness; Support functions of fuzzy sets; L_2-metric; n-dimensional random fuzzy sets; Distributions of n-dimensional random fuzzy sets; Aumann-expected value of n-dimensional fuzzy sets; Lrvc-models; Identification of parameters of lrvc-models; Least-squares estimation in lrvc-models

0. Introduction

Econometric regression models rely on observations of economic variables. The measurement of economic items is faced with the fundamental so-called problem of adequacy (cf. [12,11,24]). We have to find indicators and boundaries which represent the theoretical content of each item adequately. But many economic items, like e.g. labour force, are physically vague, that means vague in character. Therefore, the usual transformation into statistical items to intialize the process of measurement involves the risk to reduce inadequately the respective theoretical items. So the usual econometric models are idealistic in the sense that it is assumed to have crisp concepts to measure the respective economic items.

The aim of the paper is to develop regression models which relax the classical idealistic point of view. The physical vagueness of economic items should be integrated. For this purpose we want to suggest to treat economic items as linguistic variables, which leads to fuzzy data as results of measurements. This would increase the opportunities to measure economic items and might improve the consideration of the adequacy problem. As a consequence the regression models we are striving for are based on fuzzy observations of the variables. But we do not want to change the functional relationships between the variables, which means that we shall deal with crisp parameters. We shall

* Tel.: +49 681302 3169; fax: +49 681302 3551.
E-mail address: v.kraetschmer@mx.uni-saarland.de.

0165-0114/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.fss.2003.02.001

restrict ourselves to linear relationships, and we want to introduce a method of parameter estimation which extends the ordinary least-squares method.

This paper may be understood as a contribution to the topic fuzzy regression analysis. Different fuzzy regression models have been introduced to literature. Many of them treat least-squares problems but outside any stochastic environment (e.g. [5,6,28–30]). Näther and his co-workers Körner and Wünsche have investigated stochastic regression models with implemented random fuzzy sets (the extension of the notion of random variables). The earlier work of Körner and Näther is devoted to regression models with crisp exogenous variables and fuzzy parameters [13,15,25]. Recently, Wünsche and Näther also take fuzzy exogenous variables into account but consider parameters which are at least partly fuzzy [31]. Concerning our goal to find stochastic regression models which integrate the physical vagueness of the involved items without changing the initial relationships between the variables, the mentioned fuzzy regression models seem to be not suitable.

The paper is organized as follows. We shall start by providing the reader with some elements of a probability theory in sample spaces of fuzzy sets. In particular, several mathematical structures are gathered: topology, metric, σ-algebra and a semilinear structure. Afterwards we shall introduce in Section 2 the announced type of linear regression models, and we shall establish the statistical setting for the parameter estimation. Some important analytical properties of the regression functions will be stated in Section 3. In Section 4 we shall present a method to estimate the (crisp) parameter vectors of the linear regression models under consideration. It is an immediate extension of the ordinary least-squares method, which suggests to take over the name. The existence of least-squares estimators will be shown, and we shall also give a survey of some of their stochastic properties. Most of our proofs are delegated to the Appendices B, C, and the reader will be prepared for them by some auxiliary results shown in Appendix A.

1. Notations and preliminaries

Throughout this paper the Euclidean norm $\|\cdot\|$ on $\mathbb{R}^k$ is fixed, $\{e_1, \ldots, e_k\}$ stands for the standard basis of $\mathbb{R}^k$, and $\underline{0}$ for the zero vector. Any euclidean space $\mathbb{R}^l$ will be equipped with the σ-algebra $\mathcal{B}(\mathbb{R}^l)$ of Borel subsets of $\mathbb{R}^l$ generated by the standard topology. The transpose of a vector $\underline{x}$ will be denoted by $\underline{x}'$.

Furthermore, the unit interval [0, 1] will be regarded as a topological subspace of $\mathbb{R}$, endowed with the induced σ-algebra $\mathcal{B}([0,1])$ of Borel subsets. In the following we shall use the notation $L_2([0,1] \times S^0)$ for the L_2-space over the measure space $([0,1] \times S^0, \mathcal{B}([0,1]) \otimes \mathfrak{P}(S^0), \lambda^1 \otimes \lambda^{S^0})$, where $\lambda^1 \otimes \lambda^{S^0}$ indicates the product measure of λ^1, the Lebesgue–Borel measure on [0, 1], and λ^{S^0}, the uniform distribution on the powerset $\mathfrak{P}(S^0)$ of the euclidean unit sphere $S^0 := \{-1, 1\}$. The members of $L_2([0,1] \times S^0)$ are the equivalence classes $\langle f \rangle$ of real-valued functions f which are $\lambda^1 \otimes \lambda^{S^0}$-integrable of order 2. $L_2([0,1] \times S^0)$ is a vector space over $\mathbb{R}$, which is usually equipped with the euclidean scalar product $(\cdot \mid \cdot)_2$ defined by

$$(\langle f \rangle \mid \langle g \rangle)_2 := \int f g \, \mathrm{d}\lambda^1 \otimes \lambda^{S^0}.$$

The induced norm $\|\cdot\|_2$ is the well-known L_2-norm, and $L_2([0,1] \times S^0)$, equipped with $(\cdot \mid \cdot)_2$, is a separable Hilbert space.

As usual in functional analysis $L_2([0,1] \times S^0)^n$ will be endowed with the vector space structure of the respective direct sum and the scalar product $(\cdot \mid \cdot)_{n,2}$ defined by

$$((\langle f_1 \rangle, \ldots, \langle f_n \rangle) \mid (\langle g_1 \rangle, \ldots, \langle g_n \rangle))_{n,2} := \sum_{t=1}^{n} (\langle f_t \rangle \mid \langle g_t \rangle)_2.$$

The induced norm will be denoted by $\|\cdot\|_{n,2}$. $(L_2([0,1] \times S^0)^n, (\cdot \mid \cdot)_{n,2})$ is also a separable Hilbert space (cf. [7, p. 257]).

The suggestion we want to advocate is to introduce regression models which are based on data in form of fuzzy subsets of $\mathbb{R}$. In order to extend the classical single equation linear regression models, we have to link fuzzy set with probability theory. It is customary to consider the space $F_{coc}^{no}(\mathbb{R})$ which gathers all the fuzzy subsets $\widetilde{A}$ of $\mathbb{R}$ with α-cuts $[\widetilde{A}]^\alpha (\alpha \in]0,1])$ being nonvoid compact intervals. Notice that it contains the real numbers, interpreted as fuzzy subsets of $\mathbb{R}$. Every fuzzy set $\widetilde{A} \in F_{coc}^{no}(\mathbb{R})$ may be equivalently characterized by the so-called *support function*