ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 141, 1997

*Krystyna Pruska**

ON THE CLASSIFICATION OF OBSERVATIONS IN THE SWITCHING REGRESSION

Abstract. The paper discusses the method of determining the sample division indicator for the switching regression model in case of two states generating values of the explained variable, which ensures the least risk of making a mistake, understood as the expected value of relevant loss function. This paper is an attempt to take advantage of the discrimination analysis elements in the switching regression analysis.

Key words: switching regression model, discrimination analysis, loss function.

1. INTRODUCTION

The switching regression is a method of describing the dependence of a certain variable on two or more sets of variables, when the probability of determining the value of a variable explained by a defined group of explanatory variables is either known or unknown. The analyzed relations are presented by means of specific statistical models called the switching regression models.

The parameters of these models can be estimated by different methods. The maximum likelihood method is applied for this purpose most frequently. It gives consistent, asymptotically most efficient and asymptotically normal estimators of the switching regression models' parameters (see Kiefer 1978). An important fact here is having information through which the state of setting the value of the explained variable is generated, i.e. which set of the explantory variables determines this value. Very often such data are not available and decision is taken under uncertainty, on the basis of the value of some random variable, which is subjectively chosen as being adequate for performing such a role. This variable can be called on

* University of Łódź, Chair of Statistical Methods.

indicator of sample division. The way of classifying observations in the switching regression models affects the shape of the likelihood function being the basis for determining estimators of the model's parameters.

The present paper suggests the method of determining the sample division indicator for the switching regression model in case of two states generating values of the explained variable, which ensures the least risk of making a mistake, understood as the expectation of relevant loss function. This paper is an attempt to take advantage of the discriminant analysis elements in the switching regression analysis.

2. THE BASIC PROBLEM OF DISCRIMINATION

One of the problems the discriminant analysis is concerned with, is decision, basing on a random sample, on which of the two possible classes of the probability distribution the distribution of the investigated variable can be included into (see Zubrzycki 1970, p. 294–299), when the probability of which population a given sample element comes from can be either known or unknown. In this paper we shall deal with the case when this probability is known.

Let us assume that we have a random sample $X_1, ..., X_n$, selected from a population being a combination (set sum) of two populations. Let $f_1(.)$ and $f_2(.)$ denote known densities of these populations, i.e. distributions of the examined feature of given population. Let p_1 be probability that a given sample element comes from the first population, and $p_2 = 1 - p_1$ – probability that it comes from the second population. Let L_1 denote a loss resulting from classifying the element of random sample $X_1, ..., X_n$ into the second population, when in fact it comes from the first population, and L_2 – loss due to classifying a sample element into the first population, when in fact it comes from the second population (L_1 and L_2 are known values). Further let A_1 and A_2 be such sets of real numbers which are disconnected; in total they give a set of all real numbers and A_1 is the set of these values for which we conclude that a given sample element comes from the first population, and A_2 is the set of these values, for which we decide that a given sample element comes from the second population. Sets A_1 and A_2 can be defined in different ways, depending on the criterion determining the principles of decision making of classifying an observation into a specific observation. In the classical discrimination analysis sets A_1 and A_2 are determined in such a way so as to minimize the risk, i.e. the loss expected value (loss function) resulting from the way of making decision of observation classifying would be minimum.

To facilitate further considerations, let us introduce the following symbols:
r – risk function determined from pairs of sets (A_1, A_2);
L – loss function, function determining the loss we inflict, when taking the decision on classifying an observation from a sample;
C_1 – event consisting in that a given sample element comes from the first population;
C_2 – event consisting in that a given sample element comes from the second population;
D_1 – event consisting in taking the decision that a given sample element comes from the first population;
D_2 – event consisting in taking the decision that a given sample element comes from the second population;

Let us notice that for each element of the sample $X_1, ..., X_n$ the loss function L is described by the formula:

$$L = \begin{cases} L_1, & \text{when } C_1 \cap D_2 \\ L_2, & \text{when } C_2 \cap D_1 \\ 0, & \text{when } (C_1 \cap D_1) \cup (C_2 \cap D_2) \end{cases} \tag{2.1}$$

Therefore, we can determine the expectation of the loss function, that is the risk function, in the following way:

$$\begin{aligned} r = E(L) &= L_1 P(C_1 \cap D_2) + L_2 P(C_2 \cap D_1) = \\ &= L_1 P(C_1) P(D_2/C_1 + L_2 P(C_2) P(D_1/C_2), \end{aligned} \tag{2.2}$$

hence

$$r = p_1 L_1 \int_{A_2} f_1(x)dx + p_2 L_2 \int_{A_1} f_2(x)dx \tag{2.3}$$

We consider risk as the function of the sets A_1 and A_2. Hence, we search for such sets of A_1 and A_2, that the function $r(A_1, A_2)$ reached the least value. Let us notice

$$\begin{aligned} r(A_1, A_2) &= p_1 L_1 \int_{A_2} f_1(x)dx + p_1 L_1 \int_{A_1} f_1(x)dx + \int_{A_1} [p_2 L_2 f_2(x) - p_1 L_1 f_1(x)]dx = \\ &= p_1 L_1 + \int_{A_1} [p_2 L_2 f_2(x) - p_1 L_1 f_1(x)]dx \end{aligned} \tag{2.4}$$

and

$$\begin{aligned} r(A_1, A_2) &= p_2 L_2 \int_{A_1} f_2(x)dx + p_2 L_2 \int_{A_2} f_2(x)dx + \int_{A_2} [p_1 L_1 f_1(x) - p_2 L_2 f_2(x)]dx = \\ &= p_2 L_2 + \int_{A_2} [p_1 L_1 f_1(x) - p_2 L_2 f_2(x)]dx \end{aligned} \tag{2.5}$$

So we reach the following description of optimum A_1 sets A_2:

$$A_1 = \{x \in R:\ p_2 L_2 f_2(x) \leqslant p_1 L_1 f_1(x)\} \tag{2.6}$$

$$A_2 = \{x \in R:\ p_1 L_1 f_1(x) \leqslant p_2 L_2 f_2(x)\} \tag{2.7}$$

The sign of unstrict inequality occurs in formula (2.6), and strict inequality in formula (2.7). Because of the continuity of the investigated populations, distributions it is insignificant. There could occur a reverse case, which would not lead to different results of our considerations, i.e. to changes in the level of risk.

So we showed that in case when we know the probability p_1 that an observation comes from the first population, we are able to set optimum in terms of minimizing the risk function, sets A_1 and A_2 allowing to take decision of classifying a sample element to one of the possible populations.

3. THE SWITCHING REGRESSION MODEL

The switching regression model is a particular case of a statistical model with random coefficients. In this model, the coefficients can take only finite number of values. It means that the explained variable can have distribution belonging to one of several possible classes of distributions, that is, values of the variable can be determined (generated) by one of several possible states of setting. Some switching regression models are applied in the market disequilibrium analysis (see, e.g. Fair and Jaffee 1972, Fair and Kelejian 1974, Hartley and Mallela 1977, Laffont and Monfort 1979).

In this paper we shall deal with a particular case of the form of a switching regression model (see, e.g. Quandt 1972, Kiefer 1978, Charemza 1981, p. 94–87, Tomaszewicz 1985, p. 442–446, Pruska 1987):

$$y_t = \begin{cases} x'_{1t}\alpha_1 + \varepsilon_{1t} & \text{for } t \in T_1 \\ x'_{2t}\alpha_2 + \varepsilon_{2t} & \text{for } t \in T_2 \end{cases} \tag{3.1}$$

where $t = 1, ..., T$ and $T_1 \cup T_2 = \{1, 2, ..., T\}$ and $T_1 \cap T_2 = \varnothing$, when the sets of indices T_1 and T_2 can be either known or unknown. Other symbols are as follows:

y_t – variable explained by the model;

x_{1t}, x_{2t} – column vectors of the explanatory variables;

α_1, α_2 – column vectors of the model's structural parameters;

ε_{1t}, ε_{2t} – random components of the model; random variables with normal distributions with null expected values and variances σ_1^2 and σ_2^2, respectivelly, such that

$\operatorname{cov}(\varepsilon_{1t}, \varepsilon_{2t}) = 0$, $\operatorname{cov}(\varepsilon_{1t}, \varepsilon_{1\tau}) = 0$, $\operatorname{cov}(\varepsilon_{2t}, \varepsilon_{2\tau}) = 0$,

$\operatorname{cov}(\varepsilon_{1t}, \varepsilon_{2\tau}) = 0$, for $t \neq \tau$ and $t, \tau \in \{1, ..., T\}$.

The model's parameters (3.1) can be estimated by different methods. We may apply the Bayesian estimation (see e.g. Ferreira 1975, Swamy and Mehta 1975) or non-Bayesian estimation (see e.g. Fair and Jaffee 1972, Fair and Kelejian 1974, Quandt and Ramsey 1978, Schmidt 1982). If we use the maximum likelihood method (ML-estimation) we do it in two stages. The first stage consists of determining the likelihood function for a given model. The second stage is setting the point in which the function reaches its maximum. In this paper we shall only deal with the form of the likelihood function depending on the information we have on sets T_1 and T_2.

If we know sets τ_1 and τ_2 then the likelihood function for the model (3.1) is determined by the following formula (see, e.g. Goldfeld and Quandt 1972, p. 258–262, Pruska 1987, p. 21):

$$L(\alpha_1, \alpha_2, \sigma_1^2, \sigma_2^2) = (2\pi)^{-T/2}\sigma_1^{-\tau_1}\sigma_1^{-\tau_2}\exp\left\{-\frac{1}{2\sigma_1^2}\sum_{t\in T_1}(y_t - x'_{1t}\alpha_1)^2 - \right.$$
$$\left. -\frac{1}{2\sigma_2^2}\sum_{t\in T_2}(y_t - x'_{2t}\alpha_2)^2\right\} \tag{3.2}$$

where $\tau_1 = cardT_1$, $\tau_2 = cardT_2$.

If we do not know sets T_1 and T_2, then model (3.1) can be written down in the form:

$$y_t = \begin{cases} x'_{1t}\alpha_1 + \varepsilon_{1t} & \text{with probability } p_1 \\ x'_{2t}\alpha_2 + \varepsilon_{2t} & \text{with probability } p_2 = 1 - p_1 \end{cases} \tag{3.3}$$

where $0 < p_1 < 1$ and p_1 can assume either known or unknown value.

The likelihood function for the model (3.3.) is described by the formula (see, e.g Quandt 1972, Charemza 1981, p. 116, Pruska 1987, p. 24):

$$L(\alpha_1, \alpha_2, \sigma_1^2, \sigma_2^2, p_1) = \prod_{t=1}^{T}\left\{\frac{p_1}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{1}{2\sigma_1^2}(y_t - x'_{1t}\alpha_1)^2\right] + \right.$$
$$\left. + \frac{1-p_1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{1}{2\sigma_2^2}(y_t - x'_{2t}\alpha_2)^2\right]\right\} \tag{3.4}$$

In the process of estimating the parameters of the switching regression models, one can take advantage of additional information on the sample division. If we have at our disposal observations of the variable d_t for $t = 1, ..., T$, the distribution of which is in the form:

$$P(d_t = 1) = P(t \in T_1) = p_1$$
$$P(d_t = 0) = P(t \in T_2) = p_2 \tag{3.5}$$

then the proper use of these data can result in increased efficiency of estimators obtained by the likelihood method (see Kiefer 1978, 1979, Lee and Porter 1984). Variable d_t is sometimes called the sample division indicator.

In the case when observations of variable d_t are available, for the ML estimation of the model's parameters (3.3) we can use the joint density function of variables (y_t, d_t), which is described by the formula:

$$f(y_t, d_t) = d_t f_1(y_t | d_t = 1)P(d_t = 1) + (1 - d_t)f_2(y_t | d_t = 0)P(d_t = 0) \tag{3.6}$$

where f_1 and f_2 are conditional density functions of variable y_t, when, respectively, $d_t = 1$ or $d_t = 0$ (for the model (3.3), these are densities of the normal distributions, different in parameters). The likelihood function built on the formula (3.6) assumes value:

$$L(\alpha_1, \alpha_2, \sigma_1^2, \sigma_2^2, p_1) = \prod_{t=1}^{T} f(y_t, d_t) \tag{3.7}$$

where p_1 is either known or unknown value. If we know p_1, we need not estimate the parameter and then the function (3.7) depends only on $\alpha_1, \alpha_2, \sigma_1^2, \sigma_2^2$. The model's parameters estimators (3.3) obtained in the process of maximizing the likelihood function (3.7) are more efficient than the estimators obtained from the function (3.4) (see Kiefer 1979).

4. DISCRIMINATIVE CONSTRUCTION OF THE SAMPLE DIVISION INDICATOR

From the considerations presented in the works by Kiefer (1979) and Lee and Porter (1984) it follows, that having extra information on the observations classification, which is provided by the sample division indicator, results in increasing the efficiency of the ML-estimators of the switching regression models. There arises a question whether there is also a possibility to construct the sample division indicator. So far the observable variables (or their transformations) linked to the examined process described by means of the switching regression model, have been assumed as indicators. Constructing the sample division indicator is suggested in the same way as there are created sets of values of an investigated feature in the discriminant analysis, which quarantee minimum risk while taking decisions on including the sample element to one of the two possible populations. Some similarities between problems appearing in the switching regression analysis and the discriminant analysis have already been noticed in the work by Kiefer (1980). It includes a suggestion of a new method of estimating the switching regression model's parameters, alternative to the maximum likelihood method, but not to the construction of the sample division indicator. For, the model (3.1), wherein sets T_1 and T_2 are not known, let us create variable d_t of the form:

$$d_t = \begin{cases} 1 & \text{dla } t \in T_1 \\ 0 & \text{dla } t \in T_2 \end{cases} \tag{4.1}$$

where

$$T_1 = \{0 < t < T\colon\ y_t \in A_{1t}\},\ T_2 = \{0 < t < T\colon\ y_t \in A_{2t}\} \tag{4.2}$$

$$A_{1t} = \{y \in R\colon\ p_2 L_2 f_{2t}(y) \leqslant p_1 L_1 f_{1t}(y)\} \tag{4.3}$$

$$A_{2t} = \{y \in R\colon\ p_1 L_1 f_{1t}(y) \leqslant p_2 L_2 f_{2t}(y)\} \tag{4.4}$$

$$p_1 = P(t \in T_1),\ p_2 = P(t \in T_2) \tag{4.5}$$

and L_1 and L_2 are values of the loss which is inflicted, when undertaking a wrong decision (i.e. assuming element y_t determined by the first equation as generated by the second equation, or vice versa); f_{1t} and f_{2t} are densities of random variable y_t, when it is determined by the first and second equation, respectively.

Variable d_t described by the formula (4.1) can play the part of sample division indicator for the model (3.3), when probability p_1 is known and densities f_{1t} and f_{2t} are known, too. In case of the switching regression models we usually lack such information. Theorefore, the ML-estimation of these models, using the discriminative indicator of sample division can be performed only after estimating the model's parameters by the maximum likelihood method without an indicator. Then, the parameters of distributions determined by densities f_{1t} and f_{2t} and probabilities p_1, p_2 will also be estimated. Second estimation of the model's coefficients aims at increasing efficiency of their estimators. One should also notice that determining sets A_{1t} and A_{2t} allows to define for each $t \in T_1 \cup T_2$ a group of variables (factors), through which value y_t was generated.

5. FINAL REMARKS

In the paper there has been suggested a construction of the sample distribution indicator for a model of switching regression, using some elements of discrimination analysis. Due to this, the switching regression models can be used not only for describing and forecasting phenomena generated by various groups of factors, but also for determining which group of factors set a given value of the observed random variable. Furthermore, after reestimating of the model's parameters using the indicator of sample distribution, one can expect larger efficiency of estimators. To investigate the properties of these estimators we need relevant simulation experiments.

REFERENCES

Charemza W. (1981): *Ekonometryczne modele nierównowagi. Problemy specyfikacji i estymacji*, „Zeszyty Naukowe Uniwersytetu Gdańskiego", Gdańsk.

Fair R., Jaffee D. (1972): *Methods of Estimation for Markets in Disequilibrium*, „Econometrica", 40, p. 497–514.

Fair R., Kelejian H. (1974): *Methods of Estimation for Markets in Disequilibrium. A Further Study*, „Econometrica", 42, p. 177–190.

Ferreira P. (1975): *A Bayesian Analysis of a Switching Regression Model: Known Number of Regimes*, „Journal of the American Statistical Association", 70, p. 370–374.

Golfeld S., Quandt R. (1972): *Nonlinear Methods in Econometrics*, „North–Holland Publishing Company".

Hartley M. and Mallela P. (1977): *The Asymptotic Properties of a Maximum Likelihood Estimator for a Model of Markets in Disequilibrium*, „Econometrica", 45, p. 1205–1220.

Kiefer N. (1978): *Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model*, „Econometrica", 46, p. 427–434.

Kiefer N. (1979): *On the Value of Sample Separation Information*, „Econometrica", 47, p. 997–1003.

Kiefer N. (1980): *A Note on Switching Regressions and Logistic Discrimination*, „Econometrica", 48, p. 1065–1069.

Laffont J. and Monfort A. (1979): *Disequilibrium Econometrics in Dynamic Models*, „Journal of Econometrics", 11, p. 353–361.

Lee L. F., Porter R. H. (1984): *Switching Regression Models with Imperfect Sample Separation Information – with an Application on Cartel Stability*, „Econometrica", 52, p. 391–418.

Pruska K. (1987): *Zastosowanie metody największej wiarygodności i regresji przełącznikowej do estymacji ekonometrycznych modeli nierównowagi*, Doctor's dissertation, Łódź.

Quandt R. (1972): *A New Approach to Estimating Switching Regressions*, „Journal of the American Statistical Association", 67, p. 306–310.

Quandt R., Ramsey J. (1978): *Estimating Mixtures of Normal Distributions and Switching Regressions*, „Journal of the American Statistical Association", 73, p. 730–738.

Schmidt P. (1982): *An Improved Version of the Quandt–Ramsey MGF Estimator for Mixtures of Normal Distributions and Switching Regressions*, „Econometrica", 50, p. 501–516.

Swamy P., Mehta J. (1975): *Bayesian and Non–Bayesian Analysis of Switching Regression and of Random Coefficient Regression Models*, „Journal of the American Statistical Association", 70, p. 593–602.

Tomaszewicz A. (1985): *Jednorównaniowe modele ekonometryczne przy nieklasycznych założeniach*, WUŁ, Łódź.

Zubrzycki S. (1970): *Wykłady z rachunku prawdopodobieństwa i statystyki matematycznej*, PWN, Warszawa.

Krystyna Pruska

O KLASYFIKACJI OBSERWACJI W REGRESJI PRZEŁĄCZNIKOWEJ

W pracy zaproponowana jest metoda wyznaczania indykatora podziału próby dla pewnego modelu regresji przełącznikowej z dwoma stanami generującymi wartości zmiennej objaśnianej. Indykator ten zapewnia najmniejsze ryzyko popełnienia pomyłki przy klasyfikacji obserwacji rozumiane jako wartość oczekiwana odpowiedniej funkcji straty. Przy konstrukcji tego indykatora wykorzystuje się elementy analizy dyskryminacji.