FORECASTING OF REGRESSION MODEL UNDER CLASSIFICATION OF THE DEPENDENT VARIABLE

H. Ageeva
Belarusian State University
Minsk, BELARUS
e-mail: helenaageeva@gmail.com

Abstract

Regression model under classification of the dependent variable is considered. Asymptotic properties of plug-in predictive statistic are obtained.

1 Introduction

In this paper we consider a regression model with incompletely observed dependent variable: instead of its true value we observe only one of the given intervals (classes) in which the true value falls. We denote this type of distortion by classification. Classification is a special case of grouping [2].

In discriminant function analysis [3] we use previous observations to predict the class numbers for a future moment. However, in this paper we give a point prediction for the dependent variable.

2 Regression time series under classification of the dependent variable

Let

$$Y_t = F(X_t; \theta^0) + \xi_t, \ t = 1, \dots, T, \tag{1}$$

be a multiple regression time series defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where T is the sample size; $\theta^0 = (\theta^0_1, \dots, \theta^0_m)' \in \Theta \subset \mathbb{R}^m$ is the unknown regression vector parameter; $X_t = (X_{t,1}, \dots, X_{t,N})' \in \mathbf{X} \subseteq \mathbb{R}^N$ is the observed N-dimensional vector of predictors; $Y_t \in \mathbb{R}^1$ is the nonobservable dependent variable; $\xi_t \in \mathbb{R}^1$ is the normally distributed random error with mean $\mathbf{E}\{\xi_t\} = 0$ and unknown variance $0 < \mathbf{D}\{\xi_t^2\} = (\sigma^0)^2 < +\infty$; $\{\xi_t\}_{t=1}^n$ are jointly independent. The true model parameter is a composite vector-column $\delta^0 = (\theta^{0\prime}, (\sigma^0)^2)' \in \Xi \subseteq \mathbb{R}^{m+1}$.

Let the set of real numbers $\mathbb{R}$ be divided into K nonintersecting intervals $(2 \leq K < +\infty)$:

$$A_k = (a_{k-1}, a_k], \ k \in \mathbf{K} = \{1, 2, \dots, K\}, \ -\infty = a_0 < a_1 < \dots < a_K = +\infty. \tag{2}$$

This set of intervals defines classification of the dependent variable Y_t:

$$Y_t \text{ belongs to class } \nu_t \in \mathbf{K}, \text{ if } Y_t \in A_{\nu_t}. \tag{3}$$

Instead of exact values of $Y_1, \dots, Y_T$ we observe only corresponding class (interval) numbers $\nu_1, \dots, \nu_T \in \mathbf{K}$. Our aim is to construct a forecast of the dependent variable Y_{T+1} for some future predictor X_{T+1}.

3 Maximum likelihood estimator

Introduce the notation:

$$P(k; \delta, X) = \Phi\left(\frac{a_k - F(X;\theta)}{\sigma}\right) - \Phi\left(\frac{a_{k-1} - F(X;\theta)}{\sigma}\right),$$

where $k \in \mathbf{K}$, $\delta = (\theta', \sigma^2)' \in \Xi$, $X \in \mathbf{X}$, $\Phi(\cdot)$ is the standard normal distribution function. Model assumptions (1), (2), (3) determine the probability distribution of the random observations $\nu_t \in \mathbf{K}$:

$$\mathbf{P}_{X_t,\delta}\{\nu_t = k\} = \mathbf{P}_{X_t,\delta}\{Y_t \in A_k\} = P(k; \delta, X_t),\ t = 1, \ldots, T;$$

observations $\{\nu_t\}_{t=1}^n$ are jointly independent.

Lemma 1. *Under model assumptions (1), (2), (3) the log-likelihood function is*

$$l(\delta; H, \mathcal{X}) = \sum_{t=1}^{T} \ln\left(\Phi\left(\frac{a_{\nu_t} - F(X_t;\theta)}{\sigma}\right) - \Phi\left(\frac{a_{\nu_t - 1} - F(X_t;\theta)}{\sigma}\right)\right), \quad (4)$$

where $\mathcal{X} = \{X_1, \ldots, X_T\}$ is the experimental design, $H = \{\nu_1, \ldots, \nu_T\}$ is the set of classified observations.

Maximum likelihood estimator (MLE) $\hat{\delta}^T$ of the model parameter δ^0 is determined by maximization of the log-likelihood function (4):

$$\hat{\delta}^T = (\hat{\theta}^T, (\hat{\sigma}^T)^2)' : \quad l(\hat{\delta}^T; \mathcal{H}, \mathcal{X}) = \max_{\delta \in \Xi} l(\delta; \mathcal{H}, \mathcal{X}). \quad (5)$$

The following theorems present asymptotic properties of MLE $\hat{\delta}^T$ [1].

Theorem 1. *Let the following conditions hold:*

SC1. $K > 2$.

SC2. Regression coefficient space Θ is a closed bounded subset of $\mathbb{R}^m$; there are known bounds $\bar{\sigma}^2 > 0$ and $\bar{\bar{\sigma}}^2 > 0$, that $\bar{\sigma}^2 \leq (\sigma^0)^2 \leq \bar{\bar{\sigma}}^2$.

SC3. Regressors space $\mathbf{X} \subseteq \mathbb{R}^N$ is a compact space.

SC4. Function $F(X;\theta)$ is continuous on $\mathbf{X} \times \Theta$.

SC5. For any $\varepsilon > 0$ there exists $\gamma = \gamma(\varepsilon) > 0$ that the following limit expression

$$\lim_{T\to\infty} \tfrac{1}{T} \textstyle\sum_{t=1}^{T} \mathbf{I}_{\{|F(X_t;\theta^0) - F(X_t;\theta)| \geq \gamma\}} = b$$

holds for any $\theta \in \Theta$, $|\theta - \theta^0| \geq \varepsilon$, where $0 < b = b(\theta, \theta^0, \gamma, F(\cdot)) \leq 1$, $\mathbf{I}_{\{A\}}$ is the identifier of event A.

Then MLE $\hat{\delta}^T$ is strongly consistent:

$$\hat{\delta}^T \xrightarrow[T\to\infty]{\mathbf{P}=1} \delta^0.$$

Define Fisher information matrix:

$$\Gamma_T(\delta) = \sum_{t=1}^{T} \mathbf{E}_{X_t,\delta^0}\{(\nabla_\delta \ln P(\nu_t;\delta,X_t))(\nabla_\delta \ln P(\nu_t;\delta,X_t)')\}.$$

Theorem 2. *Let the following conditions hold:*

A1. MLE $\hat{\delta}^T$ is a consistent estimator of the parameter vector δ^0.

A2. For any fixed $\delta \in \Xi$ functions $F(X;\theta)$, $\frac{\partial F(X;\theta)}{\partial \theta_i}$, $\frac{\partial^2 F(X;\theta)}{\partial \theta_i \partial \theta_j}$, $\frac{\partial^3 F(X;\theta)}{\partial \theta_i \partial \theta_j \partial \theta_s}$, $i,j,s = 1,\ldots,m$, are bounded on $\mathbf{X}$;

A3. $\bar{\Gamma}_T(\delta^0) = \frac{1}{T}\Gamma_T(\delta)$ is a positive definite matrix: $\bar{\Gamma}_T(\delta^0) \succ 0$.

A4. $\varliminf\limits_{T\to\infty} \left|\bar{\Gamma}_T(\delta^0)\right| = b > 0$.

Then MLE $\hat{\delta}^T$ is asymptotically normal distributed:

$$\mathcal{L}\left\{T^{\frac{1}{2}}(\bar{\Gamma}_T(\delta^0)^{\frac{1}{2}})(\hat{\delta}^T - \delta^0)\right\} \xrightarrow[T\to\infty]{} \mathcal{N}_{m+1}(0_{m+1}, \mathbf{I_{m+1}}).$$

4 Plug-in predictive statistic

Under model assumptions (1), (2), (3) plug-in forecasting statistic is

$$\hat{Y}_{T+1} = F(X_{T+1};\hat{\theta}^T). \tag{6}$$

Let us present Fisher information matrix $\Gamma_T(\delta^0)^{-1}$ in a block form:

$$\Gamma_T(\delta^0)^{-1} = \begin{bmatrix} (\Gamma_T(\delta^0)^{-1})^{(1,1)} & (\Gamma_T(\delta^0)^{-1})^{(1,2)} \\ (\Gamma_T(\delta^0)^{-1})^{(2,1)} & (\Gamma_T(\delta^0)^{-1})^{(2,2)} \end{bmatrix},$$

where dimensions of matrices $(\Gamma_T(\delta^0)^{-1})^{(1,1)}$, $(\Gamma_T(\delta^0)^{-1})^{(1,2)}$, $(\Gamma_T(\delta^0)^{-1})^{(2,1)}$, $(\Gamma_T(\delta^0)^{-1})^{(2,2)}$ are $m\times m$, $m\times 1$, $1\times m$, 1×1 correspondingly.

Theorem 3. *Let MLE $\hat{\delta}^T$ be strongly consistent and asymptotically normal distributed estimation of δ^0 and function $F(X;\theta)$ be twice continuously differentiable with regard to θ. Then forecast (6) is asymptotically unbiased:*

$$\mathbf{E}_{X_{T+1},\delta^0}\{\hat{Y}_{T+1} - Y_{T+1}\} \xrightarrow[T\to\infty]{} 0,$$

and its mean squared risk is

$$R = \mathbf{E}_{X_{T+1},\delta^0}\{(\hat{Y}_{T+1} - Y_{T+1})^2\} \xrightarrow[T\to\infty]{}$$

$$\xrightarrow[T\to\infty]{} (\sigma^0)^2 + (\nabla_\delta F(X_{T+1};\theta^0))'(\Gamma_T(\delta^0)^{-1})^{(1,1)}(\nabla_\delta F(X_{T+1};\theta^0)).$$

5 Computer simulations

Consider regression time series:

$$Y_t = F(X_t; \theta^0) + \xi_t = \theta_1^0 X_{t,1}^{\theta_2^0} X_{t,2}^{\theta_3^0} + \xi_t,\ t = 1, \dots, T.$$

where $\theta^0 = (2.248, 0.404, 0.803)'$, $(\sigma^0)^2 = 1$. Let $K = 3$, $a_0 = -\infty$, $a_1 = 12$, $a_2 = 24$, $a_K = +\infty$ and $\{X_{t,1}, X_{t,2}\}_{t=1}^T$ be an analytical grid on $[0, 10] \times [0, 10]$. For each T we run $Q = 100$ Monte-Carlo simulations and find forecasts $\hat{Y}_{T+1}^q$, $q = 1, ...Q$, for $X_{T+1} = (11, 11)'$. We estimate mean squared risk using $\hat{R}_1 = \frac{1}{Q}\sum_{q=1}^{Q}\left(\hat{Y}_{n+1}^q - Y_{n+1}^q\right)^2$ and $\hat{R}_2 = \frac{1}{Q}\sum_{q=1}^{Q}\left((\hat{\sigma}^{T,q})^2 + (\nabla_\delta F(X_{T+1}; \hat{\theta}^{T,q}))'(\Gamma_T(\hat{\delta}^{T,q})^{-1})^{(1,1)}(\nabla_\delta F(X_{T+1}; \hat{\theta}^{T,q}))\right)$. Simulation results are presented in Figure 1. From the figure we see that mean squared risk converges to $(\sigma^0)^2 = 1$.

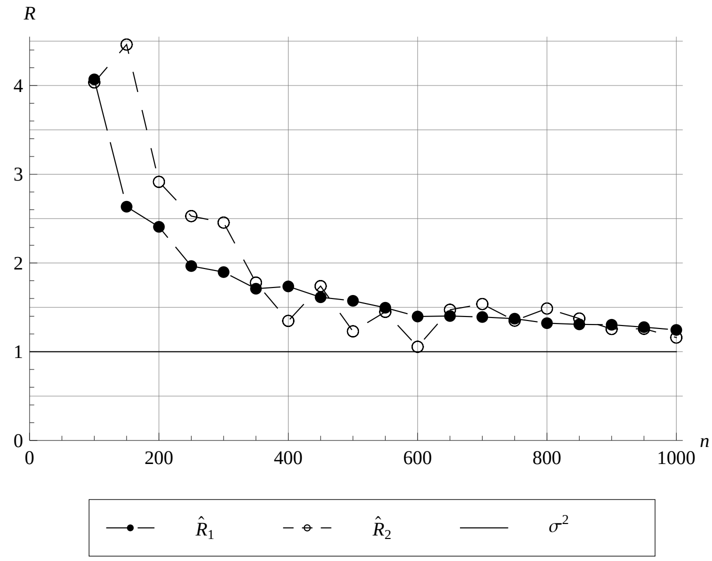

Figure 1: Estimations of squared prediction risks

References

[1] Ageeva H., Kharin Yu. (2015). ML estimation of multiple regression parameters under classification of the dependent variable. *Lithuanian Mathematical Journal*. Vol. **55**(1), pp. 48-60.

[2] Heitjan D.F. (1989). Inference from Grouped Continuous Data: A Review. *Statistical Science*. Vol. **4**(2), pp. 164-183.

[3] McLachlan G. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley.