LOGISTIC REGRESSION WITH HOMOSCEDASTIC ERRORS—A BERKSON MODEL

UDC 519.21

S. V. SHKLYAR

ABSTRACT. We consider a Berkson model of logistic regression with a single regressor and normally distributed homoscedastic errors in the regressor (the so-called Berkson model). The variance of the errors is assumed to be known. Sufficient conditions for the uniqueness of a solution of the limit estimating equation in the structural model, and sufficient conditions for the strong consistency of the maximum likelihood estimator are found in the paper.

1. INTRODUCTION

Statistical model. Consider the Berkson model of logistic regression with a single regressor. Let $X^{\text{obs}}$ be an observed value of the regressor. The true value of the regressor $X^{\text{tr}}$ given $X^{\text{obs}}$ has the normal distribution,

$$X^{\text{tr}}|X^{\text{obs}} \sim N(X^{\text{obs}}, \tau^2).$$

The response $Y$ assumes only the values 0 or 1 and let

$$P[Y=1 | X^{\text{obs}}, X^{\text{tr}}] = \frac{e^{\beta_0 + \beta_1 X^{\text{tr}}}}{1 + e^{\beta_0 + \beta_1 X^{\text{tr}}}}.$$

We further assume that the variance $\tau^2$ is the same for different observations; that is, the error is homoscedastic; the parameter $\tau^2$ is supposed to be known.

We study the structural model for which the observed values of the regressor $X_1^{\text{obs}}, X_2^{\text{obs}}, \ldots, X_N^{\text{obs}}$ are random and identically distributed. According to model (1)–(2), the triples $(X_n^{\text{obs}}, X_n^{\text{tr}}, Y_n)$, $n \geq 1$, are independent and identically distributed.

Observing the pairs $(X_n^{\text{obs}}, Y_n)$, $n=1, 2, \ldots, N$, our aim is to estimate the vector regression parameter $\beta = (\beta_0, \beta_1)^T$.

A survey of the literature. The Berkson model for the binary regression is studied in the paper [2]. In fact, [2] deals with the probit-regression, while the logistic regression is not studied there.

A model with the classical error in variables is considered in [4, 7]. Different methods of estimation are described in [4]. It is shown in [7] that the model with a single regressor is identifiable if the error is classical, if it has a normal distribution, and if the coefficient of $X$ differs from 0.

2010 Mathematics Subject Classification. Primary 62J12; Secondary 62G20.

Key words and phrases. Logistic regression, binary regression, models with errors in variables, Berkson model, regression calibration model.

The paper is based on the talk presented at the International Conference “Modern Stochastics: Theory and Applications II” held September 7–11, 2010, at Kyiv National Taras Shevchenko University and dedicated to the anniversaries of prominent Ukrainian scientists, Anatoli Skorokhod, Vladimir Korolyuk, and Igor Kovalenko.
Main result. We find conditions for the maximum likelihood estimator to be strongly consistent. We prove that if the estimator is defined as a point of maximum of the likelihood function in a compact set for which the true value of the parameter is an interior point, then the strong consistence follows if the distribution of $X^{obs}$ is not concentrated at a single point and $E|X^{obs}| < \infty$.

Under the above conditions, the model is identifiable (this is one of the assumptions for the consistence of the maximum likelihood estimator). We also prove that the same conditions imply that the limit equation $E_{\beta} S(b) = 0$ has exactly one solution in $\mathbb{R}^2$, namely $b = \beta$ (here $S(b)$ means the gradient of the logarithm of the likelihood function).

Organization of the paper. Various inequalities for the convolution of the logistic and normal densities, its antiderivative, and its derivatives are obtained in Section 2. The likelihood function and the corresponding estimating equation are given in Section 3. In Section 4, we study the limit equation $E_{\beta} S(b) = 0$. The main result, the strong consistence of the maximum likelihood estimator, is proved in Section 5.

2. The convolution of the logistic and the normal distributions

2.1. Definition of the function $L_0(x, \sigma^2)$ and its derivatives. We say that a random variable $\lambda$ has the logistic distribution if

$$P(\lambda < x) = \frac{e^x}{e^x + 1}.$$ 

Note that the density of the logistic distribution is equal to

$$e^x \frac{1}{(e^x + 1)^2}.$$ 

Consider the sum of independent random variables $\lambda + \xi$, where $\lambda$ has the logistic distribution and $\xi$ is a normal random variable; that is, $\xi \sim N(0, \sigma^2)$, where $\sigma^2 \geq 0$ (we do not exclude the case of $\sigma^2 = 0$). The distribution function of the random variable $\lambda + \xi$ is denoted by $L_0(x, \sigma^2)$. Then

$$L_0(x, \sigma^2) = P(\lambda + \xi < x) = P(\lambda < x - \xi) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(t^2/2\sigma^2)} P(\lambda < x - t) dt$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-t^2/(2\sigma^2)} \frac{e^x - t}{e^x - t + 1} dt = E e^{\frac{e^{x-\xi}}{e^{x-\xi} + 1}}.$$ 

Furthermore, let $L_1(x, \sigma^2)$ be the density of $\lambda + \xi$. For all $k \geq 1$, let

$$L_k(x, \sigma^2) = \frac{\partial^k}{\partial x^k} L_0(x, \sigma^2), \quad x \in \mathbb{R}.$$ 

Differentiating under the sign of the expectation we get

$$L_k(x, \sigma^2) = E \frac{\partial^k}{\partial x^k} \left( \frac{e^{x-\xi}}{e^{x-\xi} + 1} \right),$$

where $\xi \sim N(0, \sigma^2)$.

Note that

$$0 < L_0(x, \sigma^2) < 1, \quad L_1(x, \sigma^2) > 0$$

for all $x \in \mathbb{R}$. If $\sigma^2$ is fixed, then the function $L_0(x, \sigma^2)$ is increasing in the strict sense with respect to the argument $x$. The distribution of the sum $\xi + \lambda$ is symmetric, whence we conclude that

$$L_0(-x, \sigma^2) + L_0(x, \sigma^2) = 1.$$
Since the derivatives of the function } e^x \frac{1}{1+e^x} \text{ are bounded, the function } L_k(x, \sigma^2) \text{ is bounded, too, if } k \text{ is fixed. In particular,}

\begin{align*}
0 < L_1(x, \sigma^2) &\leq \frac{1}{4}, \\
|L_2(x, \sigma^2)| &\leq \frac{1}{6\sqrt{3}}
\end{align*}

(the inequalities are strict if } \sigma^2 > 0 \).

Next we find the derivative } \frac{\partial}{\partial t} L_0(x, t) \text{ for } t \geq 0. \text{ Let } W_t \text{ be a standard Wiener process. By Itô's formula,}

\begin{align*}
d \left( \frac{e^{x+W_t}}{e^{x+W_t} + 1} \right) &= \frac{e^{x+W_t}}{(e^{x+W_t} + 1)^2} dW_t + \frac{1}{2} \cdot \frac{e^{x+W_t} - e^{2x+2W_t}}{(e^{x+W_t} + 1)^3} dt.
\end{align*}

Passing to the expectations of both sides, we get

\begin{align*}
\mathbb{E} d \left( \frac{e^{x+W_t}}{e^{x+W_t} + 1} \right) &= \frac{1}{2} \mathbb{E} \frac{e^{x+W_t} - e^{2x+2W_t}}{(e^{x+W_t} + 1)^3} dt.
\end{align*}

In other words,

\begin{align*}
\mathbb{E} \frac{e^{x+W_{t_2}}}{e^{x+W_{t_2}} + 1} - \mathbb{E} \frac{e^{x+W_{t_1}}}{e^{x+W_{t_1}} + 1} &= \frac{1}{2} \int_{t_1}^{t_2} \mathbb{E} \frac{e^{x+W_t} - e^{2x+2W_t}}{(e^{x+W_t} + 1)^3} dt.
\end{align*}

Using equality (3) with } k = 2 \text{ we obtain

\begin{align*}
L_0(x, t_2) - L_0(x, t_1) &= \frac{1}{2} \int_{t_1}^{t_2} L_2(x, t) dt.
\end{align*}

Thus

\begin{align*}
(4) \quad \frac{\partial L_0(x, t)}{\partial t} &= \frac{1}{2} L_2(x, t).
\end{align*}

This result without proof is given in [5] with a different notation. Differentiating equality (4) with respect to } x, \text{ we get

\begin{align*}
\frac{\partial L_k(x, t)}{\partial t} &= \frac{1}{2} L_{k+2}(x, t), \quad k \geq 0.
\end{align*}

2.2. An inequality between } L_1, L_2 \text{ and } L_3.

\begin{lemma}
Let } \xi \text{ and } \eta \text{ be independent random variables and let } \xi \sim N(0, 1). \text{ Let } \zeta = \xi + \eta. \text{ Denote the density of } \zeta \text{ by } p_\zeta(z). \text{ Then

\begin{align*}
(6) \quad \frac{d^2}{dz^2} (\ln p_\zeta(z)) &= \text{Var}[\eta \mid \zeta = z] - 1,
\end{align*}

where } \text{Var}[\eta \mid \zeta = z] \text{ denotes the conditional variance of the random variable } \eta \text{ given } \zeta = z.

\begin{proof}
The density of } \zeta \text{ is equal to

\begin{align*}
p_\zeta(z) &= \mathbb{E} p_\zeta(z - \eta) = \frac{1}{\sqrt{2\pi}} \mathbb{E} e^{-(z-\eta)^2/2}.
\end{align*}

\end{proof}

Then we find the derivatives

\[
p'_c(z) = \frac{1}{\sqrt{2\pi}} \mathbb{E} \left[ (\eta - z) e^{-(z-\eta)^2/2} \right],
\]

\[
\frac{d}{dz} (\ln p_c(z)) = \frac{p'_c(z)}{p_c(z)} = \frac{\mathbb{E} \left[ (\eta - z) e^{-(z-\eta)^2/2} \right]}{\mathbb{E} e^{-(z-\eta)^2/2}} = \frac{\mathbb{E} \eta e^{-(z-\eta)^2/2}}{\mathbb{E} e^{-(z-\eta)^2/2}} - z,
\]

\[
\frac{d^2}{dz^2} (\ln p_c(z)) = \frac{\mathbb{E} \left[ \eta(\eta - z) e^{-(z-\eta)^2/2} \right] \mathbb{E} e^{-(z-\eta)^2/2} - \mathbb{E} \eta e^{-(z-\eta)^2/2} \mathbb{E} \left[ (\eta - z) e^{-(z-\eta)^2/2} \right]}{(\mathbb{E} e^{-(z-\eta)^2/2})^2} - 1
\]

\[
= \frac{\mathbb{E} \eta^2 e^{-(z-\eta)^2/2} \mathbb{E} e^{-(z-\eta)^2/2} - \left( \mathbb{E} \eta e^{-(z-\eta)^2/2} \right)^2}{(\mathbb{E} e^{-(z-\eta)^2/2})^2} - 1.
\]

If the density of \( \eta \) is denoted by \( p_\eta(y) \), then the density of the conditional distribution of \( \eta \) given \( \zeta = z \) is equal to

\[
p_{\eta|\zeta=z}(y) = \frac{p_\eta(y)p_{\zeta|\eta=y}(z)}{\int p_\eta(y_1)p_{\zeta|\eta=y_1}(z) \, dy_1} = \frac{p_\eta(y) \frac{1}{\sqrt{2\pi}} e^{-(z-y)^2/2}}{\int p_\eta(y_1) \frac{1}{\sqrt{2\pi}} e^{-(z-y_1)^2/2} \, dy_1} = \frac{p_\eta(y) e^{-(z-y)^2/2}}{\mathbb{E} e^{-(z-\eta)^2/2}}.
\]

Hence

\[
\mathbb{E}[\eta | \zeta = z] = \frac{\mathbb{E} \eta e^{-(z-\eta)^2/2}}{\mathbb{E} e^{-(z-\eta)^2/2}}, \quad \mathbb{E}[\eta^2 | \zeta = z] = \frac{\mathbb{E} \eta^2 e^{-(z-\eta)^2/2}}{\mathbb{E} e^{-(z-\eta)^2/2}}.
\]

The above results hold also in the case where the distribution of \( \eta \) is not absolutely continuous. The proof is the same but one can use the density \( dP_{\eta|\zeta=z}/dP_\eta \) with respect to the probability measure of the marginal distribution of \( \eta \) instead of the density \( p_{\eta|\zeta=z} \) with respect to the Lebesgue measure.

Thus

\[
\text{Var} [\eta | \zeta = z] = \frac{\mathbb{E} \eta^2 e^{-(z-\eta)^2/2}}{\mathbb{E} e^{-(z-\eta)^2/2}} - \left( \frac{\mathbb{E} \eta e^{-(z-\eta)^2/2}}{\mathbb{E} e^{-(z-\eta)^2/2}} \right)^2.
\]

Equalities (7) and (8) imply (6).

\[\square\]

**Corollary 2.2.** Let \( \xi \) and \( \eta \) be independent random variables such that \( \xi \) has a normal distribution; that is, \( \xi \sim N(\mu, \sigma^2) \). Consider the sum \( \zeta = \xi + \eta \) and we let \( p_c(z) \) denote the density of \( \zeta \). Then

\[
\frac{d^2}{dz^2} (\ln p_c(z)) = \frac{1}{\sigma^4} \text{Var} [\eta | \zeta = z] - \frac{1}{\sigma^2}.
\]

In particular,

\[
\frac{d^2}{dz^2} (\ln p_c(z)) \geq -\frac{1}{\sigma^2}.
\]

The inequality becomes an equality if and only if the distribution of \( \eta \) is concentrated at a single point.

**Example 2.1.** Let \( X \sim N(\mu_x, \sigma^2_x) \) and \( U \sim N(0, \sigma^2_u) \) be two independent normal random variables. If \( W = X + U \), then \( W \sim N(\mu_x, \sigma^2_w) \), where \( \sigma^2_w = \sigma^2_x + \sigma^2_u \). For all \( w \in \mathbb{R} \), we
have
\[
\frac{d^2}{dw^2} \left( \ln p_W(w) \right) = -\frac{1}{\sigma_w^2},
\]
\[
\text{Var}[X \mid W] = \frac{\sigma_u^2 \sigma_x^2}{\sigma_w^2}.
\]

This result without proof is given in [7]; the general case is considered in [1, Section 2.5]. Applying Corollary 2.2, we see that (9) becomes an equality,
\[
-\frac{1}{\sigma_w^2} = \frac{1}{\sigma_u^2} \cdot \frac{\sigma_u^2 \sigma_x^2}{\sigma_w^2} - \frac{1}{\sigma_u^2}.
\]

**Example 2.2.** Recall that, given a fixed \( \sigma^2 > 0 \), the function \( L_1(x, \sigma^2) \) is the density of the sum of two independent random variables, where one of these random variables has the logistic distribution and the other one has the normal distribution \( N(0, \sigma^2) \). By Corollary 2.2,
\[
L_3(x, \sigma^2) L_1(x, \sigma^2) - L_2(x, \sigma^2)^2 = \frac{\partial^2}{\partial x^2} \left( \ln L_1(x, \sigma^2) \right) > -\frac{1}{\sigma^2}.
\]

Since \( L_1(x, \sigma^2) > 0 \), inequality (10) can be rewritten as follows:
\[
\left( L_3(x, \sigma^2) - \frac{L_2(x, \sigma^2)^2}{L_1(x, \sigma^2)} \right) \sigma^2 + L_1(x, \sigma^2) > 0, \quad \sigma^2 \geq 0.
\]

In contrast to (10), inequality (11) holds also for the case of \( \sigma^2 = 0 \).

2.3. **Lower bound for** \( L_0(x, \sigma^2) \). We prove the following inequality:
\[
L_0(x, \sigma^2) \geq \frac{1}{2} e^{-|x|}.
\]

If \( x \geq 0 \), then
\[
L_0(x, \sigma^2) \geq \frac{1}{2} \geq \frac{1}{2} e^{-|x|}
\]
and inequality (12) is satisfied.

If \( x < 0 \), then for all \( t \geq 0 \),
\[
\frac{d}{dt} L_0(x, t) = \frac{1}{2} L_2(x, t) > 0;
\]
\[
L_0(x, \sigma^2) \geq L_0(x, 0) = \frac{e^x}{e^x + 1} > \frac{1}{2} e^{-|x|},
\]
and inequality (12) follows. Inequality (12) becomes an inequality if \( x = 0 \).

2.4. **Bounds for** \( L_1(x, \sigma^2) \) **and** \( L_2(x, \sigma^2) \).

**Lemma 2.3.** If \( x \in \mathbb{R} \) and \( \sigma^2 \geq 0 \), then
\[
|L_2(x, \sigma^2)| \leq L_0(x, \sigma^2) \leq L_0(x, \sigma^2) \left( 1 - L_0(x, \sigma^2) \right).
\]

The latter inequality becomes an equality if \( \sigma^2 = 0 \).

**Proof.** The inequality
\[
\frac{e^\xi}{(e^\xi + 1)^2} + \frac{e^\eta}{(e^\eta + 1)^2} \leq \frac{e^\xi}{(e^\xi + 1)(e^\eta + 1)} + \frac{e^\eta}{(e^\xi + 1)(e^\eta + 1)}
\]

can be checked directly. It becomes an equality if \( \xi = \eta \) (this is the only case where the above inequality becomes an equality). Then we take the expectation of both sides of
the latter inequality assuming that \( \xi \) and \( \eta \) are independent normal random variables, that is, \( \xi \sim \eta \sim N(x, \sigma^2) \):

\[
L_1(x, \sigma^2) + L_1(x, \sigma^2) \leq L_0(x, \sigma^2)(1 - L_0(x, \sigma^2)) + (1 - L_0(x, \sigma^2)) L_0(x, \sigma^2),
\]

where

\[
L_1(x, \sigma^2) \leq L_0(x, \sigma^2)(1 - L_0(x, \sigma^2)).
\]

This inequality becomes an equality if \( \xi = \eta \) almost surely; that is, if \( \sigma^2 = 0 \).

The inequality \( |L_2(x, \sigma^2)| < L_1(x, \sigma^2) \) follows by passing to the expectations in the inequality

\[
-\frac{e^{\xi}}{(1 + e^{\xi})^2} < \frac{e^{\xi} - e^{2\xi}}{(1 + e^{\xi})^3} < \frac{e^{\xi}}{(1 + e^{\xi})^2},
\]

where \( \xi \sim N(x, \sigma^2) \).

\[\square\]

3. **Maximum likelihood estimator**

Consider a regression \( Y \) with respect to \( X^{\text{obs}} \):

\[
P[Y=1 \mid X^{\text{obs}}] = \int_{-\infty}^{\infty} P[Y=1 \mid X^{\text{tr}}=x, X^{\text{obs}}] \ p_{X^{\text{tr}} \mid X^{\text{obs}}}(x) \ dx
\]

\[
= E \left[ \frac{e^{\beta_0 + \beta_1 X^{\text{tr}}}}{e^{\beta_0 + \beta_1 X^{\text{tr}}} + 1} \mid X^{\text{obs}} \right] = L_0 \left( \beta_0 + \beta_1 X^{\text{obs}}_n, \frac{\hat{\beta}_1^2 \tau^2}{1} \right).
\]

The maximum likelihood estimator is defined as a point of maximum in the set of parameters \( \Theta \) of the functional

\[
\sum_{Y_n=0}^{N} \ln \left( 1 - L_0 \left( b_0 + \beta_1 X^{\text{obs}}_n, b_1^2 \tau^2 \right) \right) + \sum_{Y_n=1}^{N} \ln L_0 \left( b_0 + \beta_1 X^{\text{obs}}_n, b_1^2 \tau^2 \right) \rightarrow \text{max}
\]

or

\[
\sum_{n=1}^{N} \ln L_0 \left( (2Y_n - 1) \left( b_0 + b_1 X^{\text{obs}}_n \right), b_1^2 \tau^2 \right) \rightarrow \text{max},
\]

where \( N \) is the number of observations and where \( b = (b_0, b_1) \) is the variable used to maximize the functional.

Differentiating the left-hand side of (13) and performing some simple algebra, we obtain the system of equations for the estimator (under the assumption that the estimator is an interior point of the parametric set):

\[
\sum_{n=1}^{N} \frac{L_0 \left( \hat{\beta}_0 + \hat{\beta}_1 X^{\text{obs}}_n, \hat{\beta}_1^2 \tau^2 \right) - Y_n}{L_0 \left( \hat{\beta}_0 + \hat{\beta}_1 X^{\text{obs}}_n, \hat{\beta}_1^2 \tau^2 \right) \left( 1 - L_0 \left( \hat{\beta}_0 + \hat{\beta}_1 X^{\text{obs}}_n, \hat{\beta}_1^2 \tau^2 \right) \right)} \frac{\partial L_0 \left( \hat{\beta}_0 + \hat{\beta}_1 X^{\text{obs}}_n, \hat{\beta}_1^2 \tau^2 \right)}{\partial \beta} = 0;
\]

that is,

\[
\begin{align*}
\sum_{n=1}^{N} \frac{(L_0 - Y_n) \cdot L_1}{L_0 \cdot (1 - L_0)} &= 0, \\
\sum_{n=1}^{N} \frac{(L_0 - Y_n) \left( L_1 X^{\text{obs}}_n + \hat{\beta}_1 \tau^2 L_2 \right)}{L_0 \cdot (1 - L_0)} &= 0,
\end{align*}
\]

where \( L_0, L_1, \) and \( L_2 \) are considered at the point \( \left( \hat{\beta}_0 + \hat{\beta}_1 X^{\text{obs}}_n, \hat{\beta}_1^2 \tau^2 \right) \). Therefore the expressions for \( L_0, L_1, \) and \( L_2 \) are different for different terms and depend on unknown variables of the system of equations.
4. Uniqueness of the solution of the limit equation

4.1. Equation $L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2) = L_0(\gamma_0 + \gamma_1 x, \gamma_1^2 \tau^2)$.

Lemma 4.1. If $\beta_0 \neq \gamma_0$ or if $\beta_1 \neq \gamma_1$, then the equation

$$L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2) = L_0(\gamma_0 + \gamma_1 x, \gamma_1^2 \tau^2)$$

has at most one solution.

Proof. If $\beta_1 = \gamma_1$, then $\beta_0 \neq \gamma_0$ and thus $\beta_0 + \beta_1 x \neq \gamma_0 + \gamma_1 x$ for all $x$. On the other hand, $\beta_1^2 \tau^2 = \gamma_1^2 \tau^2$. Since the function $L_1(z, \sigma^2)$ increases with respect to $z$ if $\sigma^2$ is fixed, equation (15) does not have solutions.

Now we consider the case of $\beta_1 \neq \gamma_1$. Without loss of generality we assume that $\beta_1 > \gamma_1$. Assume that equation (15) has a solution $x_1$. Then

$$\frac{d}{dx} L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2) > \frac{d}{dx} L_0(\gamma_0 + \gamma_1 x, \gamma_1^2 \tau^2)$$

at the point $x = x_1$.

Indeed, the function $z \mapsto L_0(z, \sigma^2)$ is a bijection $\mathbb{R} \to (0, 1)$ if $\sigma^2$ is fixed. Now we determine $\mu(\tilde{b}_1)$ from the equation

$$L_0(\mu(\tilde{b}_1), \tilde{b}_1^2 \tau^2) = L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2).$$

Note that $\mu(\beta_1) = \beta_0 + \beta_1 x_1$ and $\mu(\gamma_1) = \gamma_0 + \gamma_1 x_1$. The derivative $\mu'(\tilde{b}_1)$ exists by the implicit function theorem. The derivative is found from the equation

$$\frac{\partial}{\partial \tilde{b}_1} L_0(\mu(\tilde{b}_1), \tilde{b}_1^2 \tau^2) = 0;$$

$$L_1(\mu(\tilde{b}_1), \tilde{b}_1^2 \tau^2) \cdot \mu'(\tilde{b}_1) + \tilde{b}_1^2 \tau^2 L_2(\mu(\tilde{b}_1), \tilde{b}_1^2 \tau^2) = 0,$$

$$\mu'(\tilde{b}_1) = - \frac{\tilde{b}_1^2 \tau^2 L_2(\mu(\tilde{b}_1), \tilde{b}_1^2 \tau^2)}{L_1(\mu(\tilde{b}_1), \tilde{b}_1^2 \tau^2)}.$$

Now we determine the sign of the derivative

$$\frac{\partial}{\partial \tilde{b}_1} \left( \tilde{b}_1 L_1(\mu(\tilde{b}_1), \tilde{b}_1^2 \tau^2) \right)$$

$$= L_1(\mu(\tilde{b}_1), \tilde{b}_1^2 \tau^2) + \tilde{b}_1 L_2(\mu(\tilde{b}_1), \tilde{b}_1^2 \tau^2) \mu'(\tilde{b}_1) + \tilde{b}_1^2 \tau^2 L_3(\mu(\tilde{b}_1), \tilde{b}_1^2 \tau^2)$$

$$= L_1 - \frac{\tilde{b}_1^2 \tau^2 L_2^2}{L_1} + \tilde{b}_1^2 \tau^2 L_3 > 0.$$}

The latter inequality follows from (10) rewritten in the form

$$\frac{L_3 L_1 - L_2^2}{L_1^2} > - \frac{1}{\tilde{b}_1^2 \tau^2}.$$
This completes the proof of inequality \((16)\).

Thus the derivative of the continuous function \(L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2) - L_0(\gamma_0 + \gamma_1 x, \gamma_1^2 \tau^2)\) is positive at the points where the above function is equal to 0. We complete the proof by observing that such a function may equal 0 at most at one point.

\[\square\]

Remark 4.1. It follows from the proof of the above lemma that if \(\beta \neq \gamma\), then the function \(L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2) - L_0(\gamma_0 + \gamma_1 x, \gamma_1^2 \tau^2)\) changes the sign at those points where it equals 0. Moreover, if equation \((15)\) has a solution \(x = x_1\), then

\[\text{sign}(L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2) - L_0(\gamma_0 + \gamma_1 x, \gamma_1^2 \tau^2)) = \text{sign}(\beta_1 - \gamma_1) \text{sign}(x - x_1),\]

where \(\text{sign}(x) = -1\) for \(x < 0\), \(\text{sign}(x) = 0\) for \(x = 0\), and \(\text{sign}(x) = 1\) for \(x > 0\).

If \(\beta \neq \gamma\), then equality \((17)\) follows from the proof of Lemma 4.1. Finally, if \(\beta = \gamma\), then equality \((17)\) is trivial.

4.2. Conditional expectation of the estimating function. The system \((14)\) of equations for the maximum likelihood estimator of the regression parameter is written as follows:

\[\sum_{n=1}^{N} s(X_{n}\text{obs}, Y_n; b) = 0,\]

where \(s(X_{\text{obs}}, Y; b)\) is the elementary estimating function,

\[s(x, y; b) = \left(\frac{(L_0 - y)L_1}{L_0 \cdot (1 - L_0)} \left(\frac{(L_0 - y)(xL_1 + b_1 \tau^2 L_2)}{L_0 \cdot (1 - L_0)}\right)\right),\]

and where \(L_0, L_1,\) and \(L_2\) are considered at the point \((b_0 + b_1 x, b_1^2 \tau^2)\).

Next we consider the conditional expectation

\[s_0(X_{\text{obs}}; b) = E[s(X_{\text{obs}}, Y; b) \mid X_{\text{obs}}],\]

\[s_0(x; b) = \frac{L_0(b_0 + b_1 x, b_1^2 \tau^2) - L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2)}{L_0(b_0 + b_1 x, b_1^2 \tau^2)(1 - L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2))} \left(\frac{L_1}{xL_1 + b_1 \tau^2 L_2}\right),\]

where \(L_1\) and \(L_2\) are considered at the point \((b_0 + b_1 x, b_1^2 \tau^2)\). Throughout this section, \(\beta\) denotes the true value of the regression parameter. We do not indicate in the notation that \(s_0(x; b)\) depends on the real regression parameter \(\beta\); we rather assume that \(\beta\) is fixed. It is obvious that

\[\forall x \in \mathbb{R}: s_0(x; \beta) = 0.\]

Lemma 4.2. For every \(b \in \mathbb{R}^2, b \neq \beta\), there exists \(\theta \in \mathbb{R}^2\) such that

\[\theta^\top s_0(x_0, b) \geq 0\]

for all \(x_0 \in \mathbb{R}\). Moreover, the inequality becomes an equality for at most one point \(x_0\).

The result of the lemma can be rewritten as follows:

\[\forall b \neq \beta \exists \theta \in \mathbb{R}^2 \exists x_1 \in \mathbb{R}:\]

\[\begin{cases} \forall x_0 \in \mathbb{R}: \theta^\top s_0(x_0, b) \geq 0, \\ \forall x_0 \neq x_1: \theta^\top s_0(x_0, b) > 0. \end{cases}\]

\[\text{Proof.}\] Fix \(b \neq \beta\). The equation

\[L_0(b_0 + b_1 x, b_1^2 \tau^2) = L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2)\]

may have no solutions at all or it may have a unique solution.
First we treat the case where the equation has no solution. In this case, the values of the expression
\[ L_0(b_0 + b_1 x, b_1^2 \tau^2) - L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2) \]
are of the same sign. Let \( \theta = (1, 0)^\top \) if the above expression assumes only positive values and let \( \theta = (-1, 0)^\top \) otherwise. Since
\[
\frac{L_1(b_0 + b_1 x, b_1^2 \tau^2)}{L_0(b_0 + b_1 x, b_1^2 \tau^2)(1 - L_0(b_0 + b_1 x, b_1^2 \tau^2))} > 0
\]
(see Section 2.1), the lemma is proved.

Now we consider the case where equation (19) has a unique solution. We take this solution as \( x_1 \). According to Remark 4.1

\[
\text{sign}(L_0(b_0 + b_1 x, b_1^2 \tau^2) - L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2)) = \text{sign}(b_1 - \beta_1) \text{sign}(x - x_1)
\]
for all \( x \in \mathbb{R} \). Moreover \( b_1 \neq \beta_1 \).

Put \( \theta = (\theta_1, \theta_2)^\top \), where
\[
\theta_2 = \text{sign}(b_1 - \beta_1),
\]
\[
\theta_1 = \left( -x_1 - b_1 \tau^2 \frac{L_2(b_0 + b_1 x_1, b_1^2 \tau^2)}{L_1(b_0 + b_1 x_1, b_1^2 \tau^2)} \right) \theta_2,
\]
and consider the function
\[
f(x) = \theta_1 L_1(b_0 + b_1 x, b_1^2 \tau^2) + \theta_2 \left( xL_1(b_0 + b_1 x, b_1^2 \tau^2) + b_1 \tau^2 L_2(b_0 + b_1 x, b_1^2 \tau^2) \right).
\]
Then \( f(x_1) = 0 \) and
\[
f'(x) = \theta_1 b_1 L_2 + \theta_2 x L_2 + \theta_2 b_1 \tau^2 L_3,
\]
where \( L_1 \) and \( L_2 \) are considered at the point \((b_0 + b_1 x, b_1^2 \tau^2)\).

Now we prove that
\[
\forall x : \ (f(x) = 0 \Rightarrow \text{sign}(f'(x)) = \text{sign}(b_1 - \beta_1));
\]
that is, we prove that the values of the derivative of the function \( f(x) \) are of the same sign at the points where \( f(x) = 0 \).

Indeed, if \( f(x) = 0 \), then
\[
\theta_1 = -\frac{xL_1 + b_1 \tau^2 L_2}{L_1} \theta_2.
\]
The expression for \( f'(x) \) with this \( \theta_1 \) reduces to
\[
f'(x) = -\theta_2 \cdot \frac{xL_1 + b_1 \tau^2 L_2}{L_1} \cdot b_1 L_2 + \theta_2 L_1 + \theta_2 b_1 x L_2 + \theta_2 b_1 \tau^2 L_3
\]
\[
= \left( -\theta_2 \frac{\beta_1^2 \tau^2 L_2}{L_1} + L_1 + b_1^2 \tau^2 \right) \theta_2,
\]
whence relation (22) follows by the inequalities
\[
\frac{L_3 L_1 - L_2^2}{L_1^2} > \frac{1}{b_1^2 \tau^2}
\]
(this is, in fact, inequality (10)) and \( L_1 > 0 \).

A continuous function that satisfies (22) may equal 0 at a single point at most. Since \( f(x_1) = 0 \), we obtain
\[
\text{sign}(f(x)) = \text{sign}(b_1 - \beta_1) \text{sign}(x - x_1).
\]
According to Lemma 2.3, the coordinates of the vector 
\[ \theta^T s_0(x; b) = \frac{L_0(b_0 + b_1 x, b_1^2 \tau^2) - L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2)}{L_0(b_0 + b_1 x, b_1^2 \tau^2)(1 - L_0(b_0 + b_1 x, b_1^2 \tau^2))} \cdot f(x), \]
relations (20), (23), 0 < \( L_0(b_1 + b_1 x, b_1^2 \tau^2) < 1 \), and \( b_1 \neq \beta_1 \) imply the equality
\[ \text{sign}(\theta^T s_0(x; b)) = \text{sign}(x-x_1)^2, \]
whence the system of inequalities (18) follows. 

4.3. Uniqueness of the solution of the limit equation in the structural model.
The limit estimating function is denoted by
\[ s_\infty(b) = \mathbb{E} s(X^{\text{obs}}, Y; b) \]
provided the expectation on the right-hand side exists. Here \( s(X^{\text{obs}}, Y; b) \) is the elementary estimating function defined in Section 4.2, the joint distribution of \( X^{\text{obs}} \) and \( Y \) satisfies relations (1)–(2). The function \( s_\infty(b) \) depends on the true value of the parameter \( \beta \) and also on the distribution of \( X^{\text{obs}} \). Nevertheless we assume that this distribution and \( \beta \) are fixed.

**Lemma 4.3.** If the distribution of \( X^{\text{obs}} \) is such that \( \mathbb{E}|X^{\text{obs}}| < \infty \), then the expectation in (24) is finite.

**Proof.** According to Lemma 2.3 the coordinates of the vector \( s(X^{\text{obs}}, Y; b) \) are bounded from above by some integrable random variables, namely
\[
\begin{align*}
|s^{(1)}(X^{\text{obs}}, Y; b)| & \leq |L_0(b_0 + b_1 X^{\text{obs}}, b_1^2 \tau^2) - Y| < 1, \\
|s^{(2)}(X^{\text{obs}}, Y; b)| & \leq |L_0(b_0 + b_1 X^{\text{obs}}, b_1^2 \tau^2) - Y| \cdot (|X^{\text{obs}}| + |b_1| \tau^2) \\
& \leq |X^{\text{obs}}| + |b_1| \tau^2.
\end{align*}
\]
Therefore the expectation of the vector \( s(X^{\text{obs}}, Y; b) \) is finite. 

If the expectation of \( s(X^{\text{obs}}, Y; b) \) exists, then it is equal to
\[ \mathbb{E} s(X^{\text{obs}}, Y; b) = \mathbb{E}[\mathbb{E}[s(X^{\text{obs}}, Y; b) \mid X^{\text{obs}}]] = \mathbb{E} s_0(X^{\text{obs}}; b). \]

**Proposition 4.4.** If the distribution of \( X^{\text{obs}} \) is not concentrated at a single point and if the expectation of \( s(X^{\text{obs}}, Y; b) \) is finite, then
\[ s_\infty(b) = 0 \iff b = \beta. \]

**Proof.** If the expectation of the vector \( s(X^{\text{obs}}, Y; \beta) \) is finite, then \( \mathbb{E} s(X^{\text{obs}}, Y; \beta) = 0 \) by equality (27), since \( s_0(X^{\text{obs}}; \beta) = 0 \).

If \( b \neq \beta \), then Lemma 4.2 implies that there are a vector \( \theta \in \mathbb{R}^2 \) and a number \( x_1 \in \mathbb{R} \) such that
\[ \theta^T s_0(X^{\text{obs}}, b) \geq 0, \]
\[ \theta^T s_0(X^{\text{obs}}, b) > 0 \quad \text{if} \ X^{\text{obs}} \neq x_1. \]
Then equality (27) yields
\[ \theta^T s_\infty(b) = \mathbb{E} [\theta^T s_0(X^{\text{obs}}, b)] > 0, \]
since the distribution of \( X^{\text{obs}} \) is not concentrated at a single point \( x_1 \). Thus \( s_\infty(b) \neq 0 \).

**Corollary 4.5.** If \( \mathbb{E}|X^{\text{obs}}| < \infty \) and the distribution of \( X^{\text{obs}} \) is not concentrated at a single point, then the equation \( s_\infty(b) = 0 \) has a unique solution, namely \( b = \beta \).
5. CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATOR

In order to apply the theorem on the consistency of the maximum likelihood estimator for the vector parameter, we assume that the distribution $P_{X_{\text{obs}}}$ of the regressor $X_{\text{obs}}$ is known. The maximum likelihood estimator of the parameter $\beta$ does not depend on whether or not we know the distribution of $X_{\text{obs}}$. Then we take the product measure $\nu = P_{X_{\text{obs}}} \otimes (\delta_0 + \delta_1)$ as a majorizing measure for the distribution of the pair $(X_{\text{obs}}, Y)$. Here $\delta_0$ and $\delta_1$ are degenerate probability measures in $\{0, 1\}$ whose full mass is concentrated at the points 0 and 1, respectively. This allows one to omit the factor $p_{X_{\text{obs}}}(x)$ in the expression for the likelihood function.

**Proposition 5.1.** Assume that the set of parameters $\Theta \subset \mathbb{R}^2$ for the structural model is closed, bounded, and that the true value of the parameter $\beta$ is an interior point of the set $\Theta$. Let

$$\hat{\beta}_{\text{ML}} = \arg \max_{b \in \Theta} \sum_{n=1}^N L_0((2Y_n - 1) (b_0 + b_1 X_{\text{obs}}^n), b_1^2 \tau^2)$$

be the maximum likelihood estimator. If the distribution of $X_{\text{obs}}$ is not concentrated at a single point and if $E |X_{\text{obs}}| < \infty$, then $\hat{\beta}_{\text{ML}}$ is a strongly consistent estimator of the parameter $\beta$. In other words, $\hat{\beta}_{\text{ML}} \to \beta$ almost surely as $N \to \infty$.

**Proof.** We apply the theorem on the consistency of the maximum likelihood estimator in [6]. The likelihood function for a single observation is equal to

$$f(x, y; \beta) = L_0((2y - 1)(\beta_0 + \beta_1 x), \beta_1^2 \tau^2).$$

Now we check the conditions of the theorem mentioned above.

(c1) The model is identifiable; that is,

$$\nu(\{(x, y) \mid f(x, y; b) \neq f(x, y; b')\}) > 0$$

for all $b \neq \beta$. This condition is equivalent to the following one:

$$P(L_0(b_0 + b_1 X_{\text{obs}}, b_1^2 \tau^2) \neq L_0(\beta_0 + \beta_1 X_{\text{obs}}, \beta_1^2 \tau^2)) > 0, \quad b \neq \beta,$$

which holds in view of Lemma 4.1.

(c2) The set of parameters $\Theta$ is a compact set in $\mathbb{R}^2$ and the true value of the parameter $\beta$ is its interior point.

(c3) For all $(x, y)$, the likelihood function $f(x, y; b)$ is strictly positive and continuously differentiable with respect to $b$; the random variables

$$\ln f(X_{\text{obs}}, Y; b) \quad \text{and} \quad \frac{\partial}{\partial b} (\ln f(X_{\text{obs}}, Y; b))$$

are bounded from above by an integrable random variable.

The upper bound for the random variable $\ln f(X_{\text{obs}}, Y; b)$ follows from inequality (12):

$$-|b_0 + b_1 X_{\text{obs}}| - \ln 2 \leq \ln f(X_{\text{obs}}, Y; b) < 0,$$

while the upper bound for

$$\frac{\partial}{\partial b} (\ln f(X_{\text{obs}}, Y; b)) = -s(X_{\text{obs}}, Y; b)$$

holds in view of the inequalities 26 and 25. \qed
6. Conclusion

The strong consistence of the maximum likelihood estimator in the structural Berkson model of the logistic regression is proved in the paper.

Consider the structural model of the logistic regression with a single normal regressor and with the classical error. Making the linear change of the independent variable, this model reduces to the Berkson model [3, Section 2.2.3]. Applying Proposition 4.4 for the model with the changed variable, one can check that the logistic regression with the classical error satisfies condition (vii) of the paper [8].

Several particular cases suggest the following conjecture.

**Conjecture.** If not all $X_{n}^{\text{obs}}$ are identical, then the system of equations (14) for the maximum likelihood estimator has at most one solution in $\mathbb{R}^2$.

If this conjecture is true, then the construction of the maximum likelihood estimator becomes simpler. First, one can drop the assumption that the set of parameters $\Theta$ is bounded. Second, a stationary point of the likelihood function automatically is the point of its global maximum.

**Bibliography**


Department of Probability Theory, Statistics, and Actuarial Mathematics, Faculty for Mechanics and Mathematics, National Taras Shevchenko University, Academician Glushkov Avenue, 2, Kiev 03127, Ukraine

*E-mail address:* shklyar@mail.univ.kiev.ua

Received 07/SEP/2011

Translated by N. SEMENOV