Modern Stochastics: Theory and Applications 2 (2015) 131–146
DOI: 10.15559/15-VMSTA27

Identifiability of logistic regression with homoscedastic error: Berkson model

Sergiy Shklyar

Taras Shevchenko National University of Kyiv, Ukraine

shklyar@univ.kiev.ua (S. Shklyar)

Received: 20 May 2015, Revised: 19 June 2015, Accepted: 20 June 2015,
Published online: 7 July 2015

Abstract We consider the Berkson model of logistic regression with Gaussian and homoscedastic error in regressor. The measurement error variance can be either known or unknown. We deal with both functional and structural cases. Sufficient conditions for identifiability of regression coefficients are presented.

Conditions for identifiability of the model are studied. In the case where the error variance is known, the regression parameters are identifiable if the distribution of the observed regressor is not concentrated at a single point. In the case where the error variance is not known, the regression parameters are identifiable if the distribution of the observed regressor is not concentrated at three (or less) points.

The key analytic tools are relations between the smoothed logistic distribution function and its derivatives.

Keywords Logistic regression, binary regression, errors in variables, Berkson model, regression calibration model
2010 MSC 62J12

1 Introduction

Statistical model. Consider logistic regression with Berkson-type error in the explanatory variable. One trial is distributed as follows. X_n^{obs} is the observed (or assigned) surrogate regressor. The true regressor is $X_n = X_n^{\mathrm{obs}} + U_n$, where the error $U_n \sim N(0, \tau^2)$ is independent of X_n^{obs}. The response Y_n is a binary random variable and attains either 0 or 1 with

$$\mathsf{P}\big(Y_n{=}1 \bigm| X_n^{\mathrm{obs}}, X_n\big) = \frac{\exp(\beta_0 + \beta_1 X_n)}{1 + \exp(\beta_0 + \beta_1 X_n)}.$$

We consider both functional model and structural model. In the functional one, X_n^{obs} are nonrandom variables, and in the structural one, X_n^{obs} are i.i.d., and therefore in the latter model, $(X_n^{\mathrm{obs}}, X_n, Y_n)$ are i.i.d. random triples.

The couples $(X_n^{\mathrm{obs}}, Y_n)$, $n = 1, \dots, N$, are observed. Vector $\vec{\beta} = (\beta_0, \beta_1)^\top$ is a parameter of interest.

The error variance τ^2 can be either known or unknown, and we consider both cases. The conditions for identifiability of the model (or of the parameter $\vec{\beta}$) are presented.

Overview. Berkson models of logistic regression and probit regression were set up in Burr [1]. For probit regression, it is shown that the introduction of Berkson-type error is equivalent to augmentation of regression parameters. As a consequence, the Berkson model of probit regression is identifiable if τ^2 is known and is not identifiable if τ^2 is not known.

The identifiability of the classical model was studied by Küchenhoff [3]. He assumes that both the regressor and measurement error are normally distributed. Then univariate logistic regression is identifiable (here τ^2 can be unknown), and multiple logistic regression is not identifiable. Our results can be proved similarly to [3] if we assume that the distribution of the surrogate regressor X^{obs} has an unbounded support.

For classification of errors-in-variables regression models and various estimation methods, see the monograph by Carroll et al. [2].

Identifiability of the statistical model can be used in the proof of consistency of the estimator. For known τ^2, the strong consistency of the maximum likelihood estimator is obtained by Shklyar [4]. But if τ^2 is not known, the maximum likelihood estimator seems to be unstable (see discussion in [2] or [3]).

2 Convolution of logistic function with normal density

Consider the function

$$L_0(x, \sigma^2) = \mathsf{E}\,\frac{\exp(x-\xi)}{1+\exp(x-\xi)}, \quad \xi \sim N(0, \sigma^2),\ x \in \mathbb{R},\ \sigma^2 \ge 0, \tag{1}$$

that is, $L_0(x, 0) = \mathrm{e}^x/(1+\mathrm{e}^x)$ and

$$L_0(x, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \frac{\exp(x-t)}{1+\exp(x-t)} \mathrm{e}^{-t^2/(2\sigma^2)}\,\mathrm{d}t \quad \text{for } \sigma^2 > 0.$$

Denote the derivatives w.r.t. x

$$L_k(x, \sigma^2) = \frac{\partial^k}{\partial x^k} L_0(x, \sigma^2). \tag{2}$$

Differentiation of $L_k(x, \sigma^2)$ with respect to the second argument is described in Appendix A.

The distribution of Y_i given X_i^{obs} is

$$\mathsf{P}[Y_i = 1 \mid X_i^{\mathrm{obs}}] = \mathsf{E}[\mathsf{P}[Y_i = 1 \mid X_i^{\mathrm{obs}}, X_i] \mid X_i^{\mathrm{obs}}]$$

$$= \mathsf{E}\left[\frac{\exp(\beta_0+\beta_1 X_i)}{1+\exp(\beta_0+\beta_1 X_i)}\,\middle|\, X_i^{\mathrm{obs}}\right] = L\big(\beta_0+\beta_1 X_i^{\mathrm{obs}},\ \beta_1^2\tau^2\big) \tag{3}$$

since $[\beta_0+\beta_1 X_i \mid X_i^{\mathrm{obs}}] \sim N(\beta_0+\beta_1 X_i^{\mathrm{obs}},\ \beta_1^2\tau^2)$.

3 Identifiability when τ^2 is known

Theorem 1. *If in the functional model not all X^{obs} are equal, then the model is identifiable.*

Proof. Suppose that for two values of parameters $\vec\beta^{(1)} = (\beta_0^{(1)}, \beta_1^{(1)})$ and $\vec\beta^{(2)} = (\beta_0^{(2)}, \beta_1^{(2)})$, $\vec\beta^{(1)} \neq \vec\beta^{(2)}$, the distributions of observations are equal. Then for all $i = 1, 2, \ldots, N$,

$$\mathsf{P}_{\vec\beta^{(1)}}(Y_i = 1) = \mathsf{P}_{\vec\beta^{(2)}}(Y_i = 1),$$
$$L_0\big(\beta_0^{(1)} + \beta_1^{(1)} X_i^{\mathrm{obs}}, \big(\beta_1^{(1)}\big)^2\tau^2\big) = L_0\big(\beta_0^{(2)} + \beta_1^{(2)} X_i^{\mathrm{obs}}, \big(\beta_1^{(2)}\big)^2\tau^2\big).$$

However, by Lemma 4.1 from [4] the equation

$$L_0\big(\beta_0^{(1)} + \beta_1^{(1)} x, \big(\beta_1^{(1)}\big)^2\tau^2\big) = L_0\big(\beta_0^{(2)} + \beta_1^{(2)} x, \big(\beta_1^{(2)}\big)^2\tau^2\big)$$

has no more than one solution x. Hence, all X_i^{obs} are equal. □

By definition the degenerate distribution is the distribution concentrated at a single point. For the next theorem, see the proof of Theorem 5.1 in [4].

Theorem 2 ([4])**.** *If in the structural model the distribution of X_1^{obs} is not degenerate, then the parameter $\vec\beta$ is identifiable.*

4 Identifiability when τ^2 is unknown

For fixed σ^2, the function $L_0(x, \sigma^2)$ is a bijection $\mathbb{R} \to (0,\ 1)$. Hence, for fixed σ_1^2 and σ_2^2, the relation

$$L_0\big(y, \sigma_1^2\big) = L_0\big(x, \sigma_2^2\big) \tag{4}$$

sets the bijection $\mathbb{R} \to \mathbb{R}$; see Fig. 1.

Lemma 3. *For fixed $\sigma_1^2 \geq 0$ and $\sigma_2^2 \geq 0$, the sign of the second derivative of the implicit function* (4) *is*

$$\operatorname{sign}\left(\frac{\mathrm{d}^2 y}{\mathrm{d}x^2}\right) = \operatorname{sign}\big(\sigma_2^2 - \sigma_1^2\big)\operatorname{sign}(x).$$

Proof. Differentiating (4), we get

$$L_1\big(y, \sigma_1^2\big)\,\mathrm{d}y = L_1\big(x, \sigma_2^2\big)\,\mathrm{d}x;$$

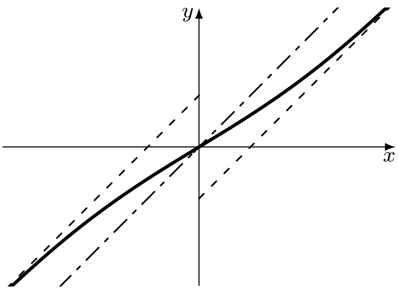

Fig. 1. The plot to equation $L_0(y,\sigma_1^2) = L_0(x,\sigma_2^2)$ for $\sigma_1^2 < \sigma_2^2$

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{L_1(x,\sigma_2^2)}{L_1(y,\sigma_1^2)}.$$

Then

$$\begin{aligned}\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} &= \frac{L_2(x,\sigma_2^2)L_1(y,\sigma_1^2) - L_1(x,\sigma_2^2)L_2(y,\sigma_1^2)\frac{\mathrm{d}y}{\mathrm{d}x}}{L_1(y,\sigma_1^2)^2} \\ &= \frac{L_2(x,\sigma_2^2)L_1(y,\sigma_1^2)^2 - L_1(x,\sigma_2^2)^2 L_2(y,\sigma_1^2)}{L_1(y,\sigma_1^2)^3} \\ &= \left(\frac{L_2(x,\sigma_2^2)}{L_1(x,\sigma_2^2)^2} - \frac{L_2(y,\sigma_1^2)}{L_1(y,\sigma_1^2)^2}\right) \cdot \frac{L_1(x,\sigma_2^2)^2}{L_1(y,\sigma_1^2)}.\end{aligned}$$

Thus,

$$\operatorname{sign}\left(\frac{\mathrm{d}^2 y}{\mathrm{d}x^2}\right) = \operatorname{sign}\left(\frac{L_2(x,\sigma_2^2)}{L_1(x,\sigma_2^2)^2} - \frac{L_2(y,\sigma_1^2)}{L_1(y,\sigma_1^2)^2}\right). \tag{5}$$

Denote by $\mu(z,\sigma^2)$ the solution to the equation $L_0(\mu,\sigma^2) = z$. Note that as $L_0(x,\sigma^2)$ is the cdf of a symmetric distribution, $\operatorname{sign}(L_0(x,\sigma^2) - 0.5) = \operatorname{sign}(x)$. Therefore, $\operatorname{sign}(\mu(z,\sigma^2)) = \operatorname{sign}(z - 0.5)$. Find the derivative

$$\frac{\mathrm{d}}{\mathrm{d}v}\left(\frac{L_2(\mu(z,v),v)}{L_1(\mu(z,v),v)^2}\right)$$

for fixed z. By the implicit function theorem,

$$\frac{\mathrm{d}\mu(z,v)}{\mathrm{d}v} = -\frac{L_2(\mu(z,v),v)}{2L_1(\mu(z,v),v)};$$

also,

$$\begin{aligned}\frac{\partial}{\partial x}\left(\frac{L_2(x,v)}{L_1(x,v)^2}\right) &= \frac{L_3(x,v)L_1(x,v) - 2L_2(x,v)^2}{L_1(x,v)^3}, \\ \frac{\partial}{\partial v}\left(\frac{L_2(x,v)}{L_1(x,v)^2}\right) &= \frac{L_4(x,v)L_1(x,v) - 2L_2(x,v)L_3(x,v)}{2L_1(x,v)^3}.\end{aligned}$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}v}\left(\frac{L_2(\mu(z,v),v)}{L_1(\mu(z,v),v)^2}\right) = -\frac{L_2}{2L_1}\cdot\frac{L_3L_1-2L_2^2}{L_1^3}+\frac{L_4L_1-2L_2L_3}{2L_1^3}$$
$$=\frac{L_4L_1^2-3L_3L_2L_1+2L_2^3}{2L_1^4},$$

where L_k are evaluated at the point $(\mu(z,v),v)$. By Lemma 10,

$$\operatorname{sign}\left(\frac{\mathrm{d}}{\mathrm{d}v}\left(\frac{L_2(\mu(z,v),v)}{L_1(\mu(z,v),v)^2}\right)\right) = \operatorname{sign}\bigl(\mu(z,v)\bigr) = \operatorname{sign}(z-0.5).$$

The function $v \mapsto \frac{L_2(\mu(z,v),v)}{L_1(\mu(z,v),v)^2}$ is monotone (it is increasing for $z > 0.5$ and decreasing for $z < 0.5$). For x and y satisfying (4),

$$x = \mu\bigl(z,\sigma_2^2\bigr) \quad \text{and} \quad y = \mu\bigl(z,\sigma_1^2\bigr)$$

with $z = L_0(y,\sigma_1^2) = L_0(x,\sigma_2^2)$; note that $\operatorname{sign}(z-0.5) = \operatorname{sign}(x)$. Then

$$\operatorname{sign}\left(\frac{L_2(x,\sigma_2^2)}{L_1(x,\sigma_2^2)^2} - \frac{L_2(y,\sigma_1^2)}{L_1(y,\sigma_1^2)^2}\right) = \operatorname{sign}\bigl(\sigma_2^2-\sigma_1^2\bigr)\operatorname{sign}(x),$$

and with (5), we can obtain the desired equality

$$\operatorname{sign}\left(\frac{\mathrm{d}^2y}{\mathrm{d}x^2}\right) = \operatorname{sign}\bigl(\sigma_2^2-\sigma_1^2\bigr)\operatorname{sign}(x). \qquad \square$$

Lemma 4. *The equation*

$$L_0\bigl(\beta_0^{(1)}+\beta_1^{(1)}x,\ \sigma_1^2\bigr) = L_0\bigl(\beta_0^{(2)}+\beta_1^{(2)}x,\ \sigma_2^2\bigr) \tag{6}$$

has no more than three solutions, unless either

$$\vec{\beta}^{(1)} = \vec{\beta}^{(2)} \quad \textit{and} \quad \sigma_1^2 = \sigma_2^2 \tag{7}$$

or

$$\beta_1^{(1)} = \beta_1^{(2)} = 0 \quad \textit{and} \quad L_0\bigl(\beta_0^{(1)},\sigma_1^2\bigr) = L_0\bigl(\beta_0^{(2)},\sigma_2^2\bigr). \tag{8}$$

In exceptional cases (7) *and* (8)*, equation* (6) *is an identity.*

Proof. The proof has the following idea: if a twice differentiable function $y(x)$ satisfies (4), then the plot of the function either is a straight line (if $\sigma_1^2 = \sigma_2^2$) or intersects any straight line at no more than three points.

Consider four cases.

Case 1. $\sigma_1^2 = \sigma_2^2$. Since the function $L_0(z,\sigma^2)$ is strictly increasing in z, Eq. (6) is equivalent to

$$\beta_0^{(1)}+\beta_1^{(1)}x = \beta_0^{(2)}+\beta_1^{(2)}x.$$

Equation (6) has only one solution if $\beta_1^{(1)} \neq \beta_1^{(2)}$; it is an identity if $\vec{\beta}^{(1)} = \vec{\beta}^{(2)}$, and it has no solutions if $\beta_1^{(1)} = \beta_1^{(2)}$ but $\beta_0^{(1)} \neq \beta_0^{(2)}$.

Case 2. $\beta_1^{(2)} = 0$ and $\beta_1^{(1)} \neq 0$. For any fixed σ^2, the function $z \mapsto L_0(z, \sigma^2)$ is a bijection $\mathbb{R} \to (0,\ 1)$. Denote the inverse function $\mu(Z, \sigma^2)$: $L_0(z, \sigma^2) = Z$ if and only if $z = \mu(Z, \sigma^2)$. Equation (6) has a unique solution

$$x = \frac{\mu(L_0(\beta_0^{(2)}, \sigma_2^2), \sigma_1^2) - \beta_0^{(1)}}{\beta_1^{(1)}}.$$

Case 3. $\beta_1^{(2)} = \beta_1^{(1)} = 0$. Neither side of (6) depends on x. Equation (6) becomes $L_0(\beta_0^{(1)}, \sigma_1^2) = L_0(\beta_0^{(2)}, \sigma_2^2)$. Equation (6) either holds for all x or does not hold for any x.

Case 4. $\sigma_1^2 \neq \sigma_2^2$ and $\beta_1^{(2)} \neq 0$. Make a linear variable substitution: denote $z_2 = \beta_0^{(2)} + \beta_1^{(2)} x$. Then Eq. (6) becomes

$$L_0\biggl(\beta_0^{(1)} + \frac{\beta_1^{(1)}}{\beta_1^{(2)}} \cdot \bigl(z_2 - \beta_0^{(2)}\bigr),\ \sigma_1^2\biggr) = L_0\bigl(z_2, \sigma_2^2\bigr). \tag{9}$$

Define the function $\mathrm{z}_1(z_2)$ from the equation

$$L_0\bigl(\mathrm{z}_1(z_2), \sigma_1^2\bigr) = L_0\bigl(z_2, \sigma_2^2\bigr).$$

The function $\mathrm{z}_1(z_2) : \mathbb{R} \to \mathbb{R}$ is implicitly defined by Eq. (4): there the equality holds if and only if $y = \mathrm{z}_1(x)$. Hence, the function $\mathrm{z}_1(z_2)$ satisfies Lemma 3. Equation (9) is equivalent to

$$\mathrm{z}_1(z_2) - \beta_0^{(1)} - \frac{\beta_1^{(1)}}{\beta_1^{(2)}} \cdot \bigl(z_2 - \beta_0^{(2)}\bigr) = 0. \tag{10}$$

By Lemma 3,

$$\begin{aligned} &\operatorname{sign}\biggl(\frac{\mathrm{d}^2}{\mathrm{d}z_2^2}\biggl(\mathrm{z}_1(z_2) - \beta_0^{(1)} - \frac{\beta_1^{(1)}}{\beta_1^{(2)}} \cdot \bigl(z_2 - \beta_0^{(2)}\bigr)\biggr)\biggr) \\ &\quad= \operatorname{sign}\biggl(\frac{\mathrm{d}^2\, \mathrm{z}_1(z_2)}{\mathrm{d}z_2^2}\biggr) = \operatorname{sign}\bigl(\sigma_2^2 - \sigma_1^2\bigr) \operatorname{sign}(z_2). \end{aligned}$$

Then the derivative of the left-hand size of (10)

$$\frac{\mathrm{d}}{\mathrm{d}z_2}\biggl(\mathrm{z}_1(z_2) - \beta_0^{(1)} - \frac{\beta_1^{(1)}}{\beta_1^{(2)}} \cdot \bigl(z_2 - \beta_0^{(2)}\bigr)\biggr) \tag{11}$$

is strictly monotone on both intervals $(-\infty,\ 0]$ and $[0,\ +\infty)$, and hence (11) attains 0 no more than at two points. Then the left-hand side of (10) has no more than three intervals of monotonicity, and Eq. (10) has no more than three solutions. Equation (6) has the same number of solutions. □

Theorem 5. *If in the functional model there are four different X^{obs}, then the parameters $\vec\beta$ and $\beta_1^2\tau^2$ are identifiable.*

Proof. Suppose that there are two sets of parameters $(\vec\beta^{(1)}, (\tau^{(1)})^2)$ and $(\vec\beta^{(2)}, (\tau^{(2)})^2)$ that for a given sample of the surrogate, the regressors $\{X_{0n},\ n = 1, \ldots, N\}$ provide the same distribution of Y_n, n=1, …, N. Then for all $n = 1, \ldots, N$,

$$\mathsf{P}_{\vec\beta^{(1)},(\tau^{(1)})^2}(Y_n = 1) = \mathsf{P}_{\vec\beta^{(2)},(\tau^{(2)})^2}(Y_n = 1);$$
$$L_0\big(\beta_0^{(1)} + \beta_1^{(1)} X_n^{\mathrm{obs}},\ (\beta_1^{(1)})^2(\tau^{(1)})^2\big) = L_0\big(\beta_0^{(2)} + \beta_1^{(2)} X_n^{\mathrm{obs}},\ (\beta_1^{(2)})^2(\tau^{(2)})^2\big).$$

The equation

$$L_0\big(\beta_0^{(1)} + \beta_1^{(1)} x,\ (\beta_1^{(1)})^2(\tau^{(1)})^2\big) = L_0\big(\beta_0^{(2)} + \beta_1^{(2)} x,\ (\beta_1^{(2)})^2(\tau^{(2)})^2\big)$$

has at least four solutions. Then by Lemma 4 either

$$\vec\beta^{(1)} = \vec\beta^{(2)} \quad \text{and} \quad (\beta_1^{(1)})^2(\tau^{(1)})^2 = (\beta_1^{(2)})^2(\tau^{(2)})^2,$$

or

$$\beta_1^{(1)} = \beta_2^{(2)} = 0 \quad \text{and} \quad L_0\big(\beta_0^{(1)},\ (\beta_1^{(1)})^2(\tau^{(1)})^2\big) = L_0\big(\beta_0^{(2)},\ (\beta_1^{(2)})^2(\tau^{(2)})^2\big). \tag{12}$$

In the latter alternative,

$$(\beta_1^{(1)})^2(\tau^{(1)})^2 = (\beta_1^{(2)})^2(\tau^{(2)})^2 = 0 \quad \text{and} \quad \beta_0^{(1)} = \beta_0^{(2)}$$

since $L_0(b_0, 0) = \frac{1}{1+\mathrm{e}^{-b_0}}$ is a strictly increasing function in b_0. □

Theorem 6. *If in the structural model the distribution of X_0 is not concentrated at three (or less) points, then the parameters $\vec\beta$ and $\beta_1^2\tau^2$ are identifiable.*

Proof. Suppose that there are two sets of parameters $(\vec\beta^{(1)}, (\tau^{(1)})^2)$ and $(\vec\beta^{(2)}, (\tau^{(2)})^2)$ for which the same bivariate distribution of $(X_1^{\mathrm{obs}}, Y_1)$ is obtained. The random variable $\mathsf{P}[Y_1 = 1 \mid X_1^{\mathrm{obs}}]$ satisfies Eq. (3) almost surely for each set of parameters. Hence, the equality

$$L_0\big(\beta_0^{(1)} + \beta_1^{(1)} X_1^{\mathrm{obs}},\ (\beta_1^{(1)})^2(\tau^{(1)})^2\big) = L_0\big(\beta_0^{(2)} + \beta_1^{(2)} X_1^{\mathrm{obs}},\ (\beta_1^{(2)})^2(\tau^{(2)})^2\big)$$

holds almost surely. The rest of the proof is the same as in Theorem 5. □

A Differentiation of $L_k(x, \sigma^2)$

Consider the sum of two independent random variables $\zeta = \lambda + \xi$, where λ has the logistic distribution

$$\mathsf{P}(\lambda \le x) = \frac{\exp(x)}{1 + \exp(x)}, \qquad x \in \mathbb{R},$$

and $\xi \sim N(0, \sigma^2)$. We allow $\sigma^2 = 0$, and then $\xi = 0$ almost surely.

The function $L_0(x,\sigma^2)$ defined in (1) is the cdf of ζ, and the function $L_1(x,\sigma^2)$ defined in (2) is the pdf of ζ.

The partial derivatives of $L_k(x,v)$ are

$$\frac{\partial}{\partial x} L_k(x,v) = L_{k+1}(x,v), \qquad \frac{\partial}{\partial v} L_k(x,v) = \frac{1}{2} L_{k+2}(x,v);$$

see the proof in [4, Section 2]. The functions $L_k(x,v)$ are infinitely differentiable and bounded on $\mathbb{R}\times[0,+\infty)$.

Since the distribution of ζ is symmetric,

$$L_k\big(-x,\sigma^2\big) = (-1)^{k-1} L_k\big(x,\sigma^2\big), \quad k \ge 1,$$

that is, $L_1(x,\sigma^2)$ and $L_3(x,\sigma^2)$ are even functions in x, and $L_2(x,\sigma^2)$ and $L_4(x,\sigma^2)$ are odd functions in x.

B The key inequality

The next lemma is similar to Lemma 2.1 in [4]. Hence, the proof is brief; see [4] for details.

Lemma 7. *Let ξ and η be two independent random variables, where $\xi \sim N(0,1)$. Denote $\zeta = \xi + \eta$ and let $p_\zeta(z)$ be the pdf of ζ. Then*

$$\frac{\mathrm{d}^3}{\mathrm{d}z^3}\big(\ln p_\zeta(z)\big) = \mu_3[\eta \mid \zeta{=}z],$$

where $\mu_3[\eta \mid \zeta{=}z]$ is the third conditional central moment,

$$\mu_3[\eta \mid \zeta{=}z] = \mathsf{E}\Big[\big(\eta - \mathsf{E}[\eta \mid \zeta{=}z]\big)^3 \;\Big|\; \zeta{=}z\Big].$$

Proof. We have

$$p_\zeta(z) = \mathsf{E}\, p_\xi(z-\eta) = \frac{1}{\sqrt{2\pi}}\, \mathsf{E}\, \mathrm{e}^{-\frac{1}{2}(z-\eta)^2}.$$

Then

$$p'_\zeta(z) = \frac{1}{\sqrt{2\pi}}\, \mathsf{E}\Big[(\eta - z)\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\Big],$$

$$\frac{\mathrm{d}}{\mathrm{d}z}\big(\ln p_\zeta(z)\big) = \frac{p'_\zeta(z)}{p_\zeta(z)} = \frac{\mathsf{E}[(\eta - z)\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}]}{\mathsf{E}\, \mathrm{e}^{-\frac{1}{2}(z-\eta)^2}} = \frac{\mathsf{E}\, \eta \mathrm{e}^{-\frac{1}{2}(z-\eta)^2}}{\mathsf{E}\, \mathrm{e}^{-\frac{1}{2}(z-\eta)^2}} - z,$$

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}\big(\ln p_\zeta(z)\big) = \frac{\mathsf{E}\, \eta^2 \mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\, \mathsf{E}\, \mathrm{e}^{-\frac{1}{2}(z-\eta)^2} - (\mathsf{E}\, \eta \mathrm{e}^{-\frac{1}{2}(z-\eta)^2})^2}{(\mathsf{E}\, \mathrm{e}^{-\frac{1}{2}(z-\eta)^2})^2} - 1,$$

$$\begin{aligned}\frac{\mathrm{d}^3}{\mathrm{d}z^3}\bigl(\ln p_\zeta(z)\bigr) &= \bigl(\mathsf{E}\,\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\bigr)^{-3}\\ &\quad\times\Bigl(\mathsf{E}\bigl[\eta^2(\eta-z)\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\bigr]\bigl(\mathsf{E}\,\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\bigr)^2\\ &\quad+\mathsf{E}\,\eta^2\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\,\mathsf{E}\bigl[(\eta-z)\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\bigr]\,\mathsf{E}\,\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\\ &\quad-2\,\mathsf{E}\bigl[\eta(\eta-z)\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\bigr]\,\mathsf{E}\,\eta\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\,\mathsf{E}\,\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\\ &\quad-2\,\mathsf{E}\,\eta^2\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\,\mathsf{E}\,\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\,\mathsf{E}\bigl[(\eta-z)\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\bigr]\\ &\quad+2\bigl(\mathsf{E}\,\eta\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\bigr)^2\,\mathsf{E}\bigl[(\eta-z)\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\bigr]\Bigr)\\ &=\bigl(\mathsf{E}\,\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\bigr)^{-3}\times\Bigl(\mathsf{E}\,\eta^3\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\bigl(\mathsf{E}\,\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\bigr)^2\\ &\quad-3\,\mathsf{E}\,\eta^2\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\,\mathsf{E}\,\eta\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\,\mathsf{E}\,\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}+2\bigl(\mathsf{E}\,\eta\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}\bigr)^3\Bigr).\end{aligned}\tag{13}$$

If η has a pdf, the conditional pdf of η given $\zeta=z$ is equal to

$$p_{\eta\mid\zeta=z}(y)=\frac{p_\eta(y)\mathrm{e}^{-\frac{1}{2}(z-y)^2}}{\mathsf{E}\,\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}};$$

otherwise, we can use the conditional density of η w.r.t. marginal density

$$\frac{\mathrm{d}\,\mathrm{cdf}_{\eta\mid\zeta=z}(y)}{\mathrm{d}\,\mathrm{cdf}_\eta(y)}=\frac{\mathrm{e}^{-\frac{1}{2}(z-y)^2}}{\mathsf{E}\,\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}}.$$

Anyway, the conditional moments of η given $\zeta=z$ are equal to

$$\mathsf{E}\bigl[\eta^k\mid\zeta=z\bigr]=\frac{\mathsf{E}\,\eta^k\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}}{\mathsf{E}\,\mathrm{e}^{-\frac{1}{2}(z-\eta)^2}}.\tag{14}$$

From (13) and (14) it follows that

$$\begin{aligned}\frac{\mathrm{d}^3}{\mathrm{d}z^3}\bigl(\ln p_\zeta(z)\bigr)&=\mathsf{E}\bigl[\eta^3\mid\zeta=z\bigr]-3\,\mathsf{E}\bigl[\eta^2\mid\zeta=z\bigr]\,\mathsf{E}[\eta\mid\zeta=z]+2\bigl(\mathsf{E}[\eta\mid\zeta=z]\bigr)^3\\ &=\mu_3[\eta\mid\zeta=z].\end{aligned}$$ □

Corollary 8. *Let ξ and η be independent random variables such that $\xi\sim N(\mu,\sigma^2)$. Denote $\zeta=\xi+\eta$, and denote the pdf of ζ by $p_\zeta(z)$. Then*

$$\frac{\mathrm{d}^3}{\mathrm{d}z^3}\bigl(\ln p_\zeta(z)\bigr)=\frac{1}{\sigma^6}\,\mu_3[\eta\mid\zeta=z].$$

Lemma 9. *Assume that the distribution of a random variable X satisfies the following conditions:*

1) *X has a continuously differentiable density $p_X(x)$.*
2) *X is unimodal in the following sense: there exists a mode $M\in\mathbb{R}$ such that for all $x\in\mathbb{R}$, we have the equality* $\operatorname{sign}(p'_X(x))=\operatorname{sign}(M-x)$.

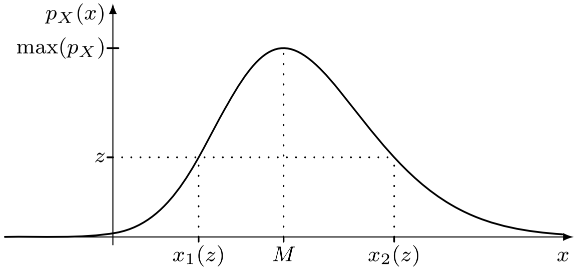

Fig. 2. To proof of Lemma 9, part 1). Sample $p_X(x)$ and definition of $x_1(z)$ and $x_2(z)$

3) *Whenever* $x_1 < M < x_2$ *and* $p_X(x_1) = p_X(x_2)$, *then* $p_X(x_1) > -p_X(x_2)$.

4) $\mathsf{E}\,|X|^3 < \infty$.

Then $\mu_3(X) := \mathsf{E}(X - \mathsf{E}\,X)^3 > 0$.

Proof. 1) $\mathsf{E}\,X > M$. Denote by $x_1(z)$ and $x_2(z)$ the solutions to the equation $p_X(x) = z$ (see Fig. 2):

$$\begin{aligned} x_1(z) < M < x_2(z) \quad & \text{if } 0 < z < \max(p_X); \\ x_1(z) = M = x_2(z) \quad & \text{if } z = \max(p_X); \\ p_X\big(x_1(z)\big) = p_X\big(x_2(z)\big) = z \quad & \text{if } 0 < z \le \max(p_X). \end{aligned}$$

Represent the expectation as a double integral and change the order of integration:

$$\begin{aligned} \mathsf{E}\,X &= M + \int_{-\infty}^{\infty} (x - M)\, p_X(x)\,\mathrm{d}x \\ &= M + \iint\limits_{\{(x,z)\,|\,0 \le z \le p_X(x)\}} (x - M)\,\mathrm{d}x\,\mathrm{d}z \\ &= M + \int_0^{\max(p_X)} \bigg(\int_{x_1(z)}^{x_2(z)} (x - M)\,\mathrm{d}x \bigg) \mathrm{d}z \\ &= M + \int_0^{\max(p_X)} \frac{(x_2(z) - M)^2 - (M - x_1(z))^2}{2}\,\mathrm{d}z. \end{aligned} \tag{15}$$

For all $x_2 > M$, by the implicit function theorem,

$$\frac{\mathrm{d}}{\mathrm{d}x_2} x_1\big(p_X(x_2)\big) = \frac{p'_X(x_2)}{p'_X(x_1(p_X(x_2)))} > -1$$

because $p_X(x_1(p_X(x_2))) = p_X(x_2)$ implies $p'_X(x_1(p_X(x_2))) > -p'_X(x_2) > 0$. Note that $x_1(p_X(M)) = M$. By the Lagrange theorem,

$$x_1\big(p_X(x_2)\big) = M + (x_2 - M) \cdot \frac{\mathrm{d}}{\mathrm{d}x_3} x_1\big(p_X(x_3)\big)\Big|_{x_3 = M + (x_2 - M)\theta}$$

for some $\theta \in (0, 1)$;

$$x_1\big(p_X(x_2)\big) > M - (x_2 - M) \quad \text{for } x_2 > M;$$

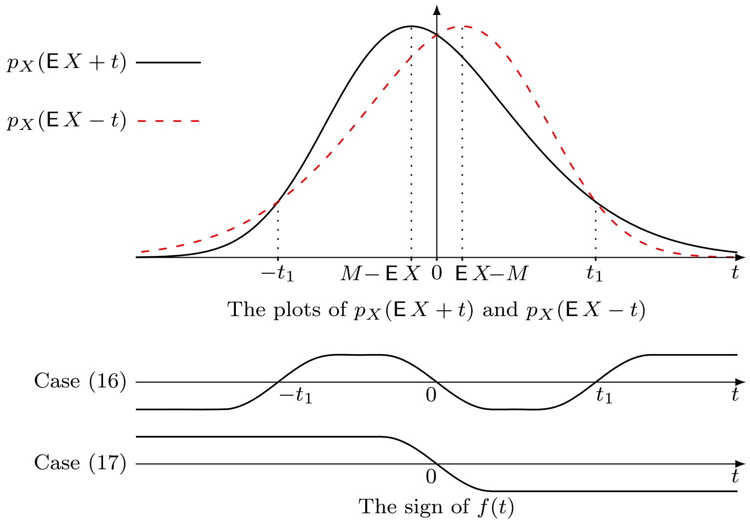

Fig. 3. To proof of Lemma 9, part 2)

$$\begin{aligned} x_1(z) &> M - \big(x_2(z) - M\big) \quad \text{for } 0 < z < \max(p_X);\\ x_2(z) - M &> M - x_1(z) > 0;\\ \frac{(x_2(z)-M)^2}{2} &> \frac{(M - x_1(z))^2}{2}; \end{aligned}$$

the last integrand in (15) is positive, and then (15) implies $\mathsf{E}\,X > M$.

2) Consider the function

$$f(t) = p_X(\mathsf{E}\,X + t) - p_X(\mathsf{E}\,X - t),$$

which is odd and strictly decreasing on the interval $[-(\mathsf{E}\,X - M),\ \mathsf{E}\,X - M]$. Therefore, $f(t)$ attains 0 only once on this interval, that is, at the point 0 (see Fig. 3).

If $t > \mathsf{E}\,X - M$ (more generally, $|t| > \mathsf{E}\,X - M$) and $f(t) = 0$, then $f'(t) = p'_X(\mathsf{E}\,X + t) + p'_X(\mathsf{E}\,X - t) > 0$ by condition 3) of Lemma 9. Therefore, $f(t)$ can attain 0 only once on $(\mathsf{E}\,X - M,\ +\infty)$, and if it attains 0 (say, at a point $t_1 > \mathsf{E}\,X - M > 0$), it is increasing in the neighborhood of t_1.

Hence, there may be two cases of sign changing of $f(t)$ (Fig. 3). Either

$$\exists t_1 > 0\ \forall x \in \mathbb{R}\ :\ \operatorname{sign}\big(f(t)\big) = \operatorname{sign}(t)\operatorname{sign}\big(|t| - t_1\big), \tag{16}$$

or

$$\forall x \in \mathbb{R}\ :\ \operatorname{sign}\big(f(t)\big) = -\operatorname{sign}(t). \tag{17}$$

3) We have

$$\begin{aligned} 0 = \mathsf{E}[X - \mathsf{E}\,X] &= \int_{-\infty}^{\infty} (x - \mathsf{E}\,X)\, p_X(x)\,\mathrm{d}x\\ &= \int_{-\infty}^{\infty} t\, p_X(\mathsf{E}\,X + t)\,\mathrm{d}t \end{aligned}$$

$$= \int_0^\infty t\, p_X(\mathsf{E}\,X + t)\,\mathrm{d}t + \int_0^\infty (-t)\, p_X(\mathsf{E}\,X - t)\,\mathrm{d}t$$
$$= \int_0^\infty t\, f(t)\,\mathrm{d}t, \tag{18}$$

where $f(t)$ is defined in the second part of the proof.

Note that the case (17) is impossible because otherwise the last integrand in (18) would be negative and thus the integral could not be equal to 0.

4) Similarly to (18),

$$\mathsf{E}(X - \mathsf{E}\,X)^3 = \int_0^\infty t^3 f(t)\,\mathrm{d}t.$$

Subtract t_1^2 times Eq. (18), where t_1 comes from (16):

$$\mathsf{E}(X - \mathsf{E}\,X)^3 = \int_0^\infty t\big(t^2 - t_1^2\big) f(t)\,\mathrm{d}t.$$

The integrand is positive for $t > 0, t \neq t_1$, and hence $\mu_3[X] = \mathsf{E}(X - \mathsf{E}\,X)^3 > 0$. □

Lemma 10. *For all $x \in \mathbb{R}$ and $\sigma^2 \geq 0$,*

$$\operatorname{sign}\big(L_4(x,\sigma^2)L_1(x,\sigma^2)^2 - 3L_3(x,\sigma^2)L_2(x,\sigma^2)L_1(x,\sigma^2) + 2L_2(x,\sigma^2)^3\big)$$
$$= \operatorname{sign}(x).$$

Lemma 11 is needed to prove Lemma 10. The notation $F(y)$ and y_0 is common for Lemmas 10 and 11.

For fixed $x > 0$ and σ^2, consider the function

$$F(y) = \ln\left(\frac{\mathrm{e}^y}{(\mathrm{e}^y + 1)^2}\right) - \frac{(y - x)^2}{2\sigma^2}. \tag{19}$$

Its derivative

$$F'(y) = 1 - 2\frac{\mathrm{e}^y}{\mathrm{e}^y + 1} - \frac{y - x}{\sigma^2}$$

is strictly decreasing, and

$$\lim_{y \to -\infty} F'(y) = +\infty, \qquad \lim_{y \to +\infty} F'(y) = -\infty.$$

Hence, $F'(y)$ attains 0 at a unique point. Denote this point by y_0, and then

$$\operatorname{sign}\big(F'(y)\big) = -\operatorname{sign}(y - y_0). \tag{20}$$

Lemma 11. *For the function $F(y)$ defined in (19), for y_0 satisfying (20), and for y_3 and y_4 such that $F'(y_3) + F'(y_4) = 0$ and $y_3 < y_4$, we have the following inequalities:*

1) $y_3 < y_0 < y_4$ *and* $F'(y_3) = -F'(y_4) > 0$.

2) $y_3 + y_4 > 0$.

3) $F''(y_3) < F''(y_4) < 0$.

4) $F(y_3) > F(y_4)$.

Proof. 1) *The inequality* $y_3 < y_0 < y_4$ is a consequence of (20), and (20) implies $F'(y_3) > 0$.

2) $y_3 + y_4 > 0$. For all $y \in \mathbb{R}$,

$$F'(y) + F'(-y) = \frac{2x}{\sigma^2} > 0.$$

Since $F'(y_3) + F'(-y_3) > 0$ and $F'(y_3) + F'(y_4) = 0$, we have $F'(-y_3) > F'(y_4)$, and then $-y_3 < y_4$ because the derivative $F'(y)$ is decreasing.

3) $F''(y_3) < F''(y_4) < 0$. The second derivative

$$F''(y) = \frac{-2\mathrm{e}^y}{(\mathrm{e}^y + 1)^2} - \frac{1}{\sigma^2}$$

is an even function strictly increasing on $[0, +\infty)$ and attaining only negative values.

The inequalities $y_3 < y_4$ and $y_3 + y_4 > 0$ can be rewritten as $|y_3| < y_4$, and then

$$F''(y_3) = F''\big(|y_3|\big) < F''(y_4) < 0.$$

4) $F(x_3) > F(x_4)$. Consider the inverse function

$$\big(F'\big)^{-1}(t), \quad t \in \mathbb{R}.$$

Its derivative is

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\big(F'\big)^{-1}(t)\Big) = \frac{1}{F''((F')^{-1}(t))} < 0.$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(F\big(\big(F'\big)^{-1}(t)\big)\Big) = \frac{F'((F')^{-1}(t))}{F''((F')^{-1}(t))} = \frac{t}{F''((F')^{-1}(t))};$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(F\big(\big(F'\big)^{-1}(t)\big) - F\big(\big(F'\big)^{-1}(-t)\big)\Big) = \frac{t}{F''((F')^{-1}(t))} + \frac{-t}{F''((F')^{-1}(-t))}.$$

Apply already proven part 3) of Lemma 11. If $t > 0$, then $(F')^{-1}(t) < (F')^{-1}(-t)$ (because $(F')^{-1}(t)$ is a decreasing function) and $F'((F')^{-1}(t)) + F'((F')^{-1}(-t)) = t - t = 0$. Then by part 3)

$$F''\big(\big(F'\big)^{-1}(t)\big) < F''\big(\big(F'\big)^{-1}(-t)\big) < 0, \quad t > 0.$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(F\big(\big(F'\big)^{-1}(t)\big) - F\big(\big(F'\big)^{-1}(-t)\big)\Big) > 0, \quad t > 0.$$

Note that

$$F\big(\big(F'\big)^{-1}(0)\big) - F\big(\big(F'\big)^{-1}(-0)\big) = 0.$$

By the Lagrange theorem, for $t > 0$,

$$F\big(\big(F'\big)^{-1}(t)\big) - F\big(\big(F'\big)^{-1}(-t)\big) = t \cdot \frac{\mathrm{d}}{\mathrm{d}t_1}\Big(F\big(\big(F'\big)^{-1}(t_1)\big) - F\big(\big(F'\big)^{-1}(-t_1)\big)\Big) > 0, \tag{21}$$

where the derivative is taken at some point $t_1 \in (0, t)$.

Substituting $t = F'(y_3) > 0$ (then $-t = F'(y_4)$), we obtain $F(y_3) - F(y_4) > 0$. □

Proof of Lemma 10. *Case 1. $x > 0$ and $\sigma^2 > 0$.* Recall that for fixed σ^2, $L_1(x, \sigma^2)$ is the pdf of $\eta + \xi$, where η and ξ are independent variables, $\mathsf{P}(\eta < y) = \frac{e^y}{e^y+1}$ and $\xi \sim N(0, \sigma^2)$ (see Appendix A). By Corollary 8,

$$\frac{d^3}{dx^3}\big(\ln L_1\big(x, \sigma^2\big)\big) = \frac{1}{\sigma^6}\mu_3[\eta \mid \eta+\xi=x], \tag{22}$$

but

$$\frac{d^3}{dx^3}\big(\ln L_1\big(x, \sigma^2\big)\big) = \frac{L_4 L_1^2 - 3L_3L_2L_1 + 2L_2^3}{L_1^3}, \tag{23}$$

where L_k are evaluated at the point (x, σ^2). Since $L_1(x, \sigma^2) > 0$, we have to prove that $\mu_3[\eta \mid \eta+\xi=x] > 0$. Therefore, we apply Lemma 9.

The pdf of the conditional distribution of η given $\eta + \xi = x$ is equal to

$$p_{\eta|\eta+\xi=x}(y) = \frac{1}{\mathsf{E}\, e^{-\frac{(\eta-x)^2}{2\sigma^2}}} \cdot \frac{e^y}{(1+e^y)^2} e^{-\frac{(y-x)^2}{2\sigma^2}}.$$

The pdf $p_{\eta|\eta+\xi=x}(y)$ is continuously differentiable. The conditional distribution has a finite kth moment because $y^k e^{-\frac{(y-x)^2}{2\sigma^2}}$ is bounded for any $k \in \mathbb{N}$. Hence, conditions 1) and 4) of Lemma 9 are satisfied.

Evaluate

$$\ln p_{\eta|\eta+\xi=x}(y) = \ln\left(\frac{e^y}{(e^y+1)^2}\right) - \frac{y-x}{2\sigma^2} - \ln\big(\mathsf{E}\, e^{-\frac{(\eta-x)^2}{2\sigma^2}}\big) = F(y) + C,$$

where the function $F(y)$ is defined in (19), and $C = -\ln(\mathsf{E}\exp(-\frac{(\eta-x)^2}{2\sigma^2}))$ depends only on x and σ^2 and does not depend on y.

We check condition 2) of Lemma 9:

$$\begin{aligned} p_{\eta|\eta+\xi=x}(y) &= e^{F(y)+C}; \\ \frac{d}{dy} p_{\eta|\eta+\xi=x}(y) &= F'(y) e^{F(y)+C}; \\ \operatorname{sign}\left(\frac{d}{dy} p_{\eta|\eta+\xi=x}(y)\right) &= \operatorname{sign}\big(F'(y)\big) = -\operatorname{sign}(y - y_0), \end{aligned} \tag{24}$$

and condition 2) holds with $M = y_0$, where y_0 is defined just above (20).

Now check condition of 3) of Lemma 9. The proof is illustrated by Fig. 4. Assume that $p_{\eta|\eta+\xi=x}(y_1) = p_{\eta|\eta+\xi=x}(y_2)$ and $y_1 < y_0 < y_2$. Then $F(y_1) = F(y_2)$.

Denote

$$y_4 = \big(F'\big)^{-1}\big(-F'(y_1)\big).$$

Then $F'(y_1) + F'(y_4) = F'(y_1) - F'(y_1) = 0$, and by (20), as $y_1 < y_0$, we have $F'(y_1) > 0$, $F'(y_4) < 0$, $y_4 > y_0 > y_1$. By Lemma 11, $F(y_1) > F(y_4)$.

Hence, $F(y_2) = F(y_1) > F(y_4)$. Because the function $F(y)$ is decreasing on $(y_0, +\infty)$ (see (20)), we have $y_2 < y_4$. Since the function $F'(y)$ is decreasing,

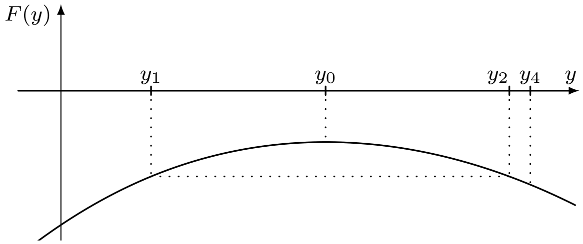

Fig. 4. To proof of Lemma 10. Checking condition 3) of Lemma 9

$F'(y_2) > F'(y_4) = -F'(y_1)$, which implies $F'(y_1) + F'(y_2) > 0$. By (24) we have $p'_{\eta|\eta+\xi=x}(y_1) + p'_{\eta|\eta+\xi=x}(y_2) > 0$.

All the conditions of Lemma 9 are satisfied. By Lemma 9, $\mu_3[\eta \mid \eta+\xi=x] > 0$, and by (22)–(23),

$$L_4(x,\sigma^2)L_1(x,\sigma^2)^2 - 3L_3(x,\sigma^2)L_2(x,\sigma^2)L_1(x,\sigma^2) + 2L_2(x,\sigma^2) > 0 \quad (25)$$

for all $x > 0$ and $\sigma^2 > 0$.

Case 2. $x \le 0$ and $\sigma^2 > 0$. The distribution of $\eta + \xi$ is symmetric. Hence, $L_1(x,\sigma^2)$ and $L_3(x,\sigma^2)$ are even functions in x, and $L_2(x,\sigma^2)$ and $L_4(x,\sigma^2)$ are odd functions in x. Then

$$L_4(x,\sigma^2)L_1(x,\sigma^2)^2 - 3L_3(x,\sigma^2)L_2(x,\sigma^2)L_1(x,\sigma^2) + 2L_2(x,\sigma^2)^3$$

is an odd function in x. It is equal to 0 for $x = 0$, and it is negative for $x < 0$ by Case 1; see (25).

Case 3. $\sigma^2 = 0$. The function $L_1(x,0)$ is the pdf of the logistic distribution, and $L_{k+1}(x,0)$ is its kth derivative:

$$L_1(x,0) = \frac{e^x}{(1+e^x)^2}; \qquad L_2(x,0) = \frac{e^x(1-e^x)}{(1+e^x)^3};$$
$$L_3(x,0) = \frac{e^x}{(1+e^x)^4}\left(1 - 4e^x + e^{2x}\right);$$
$$L_4(x,0) = \frac{e^x(1-e^x)}{(1+e^x)^5}\left(1 - 10e^x + e^{2x}\right).$$

Then

$$L_4L_1^2 - 3L_3L_2L_1 + 2L_2^3 = \frac{e^{3x}(1-e^x)}{(1+e^x)^9}\left(-2e^x\right);$$
$$\operatorname{sign}\left(L_4L_1^2 - 3L_3L_2L_1 + 2L_2^3\right) = \operatorname{sign}(x),$$

where L_k are evaluated at the point $(x, 0)$.

Lemma 10 is proven. □

References

[1] Burr, D.: On errors-in-variables in binary regression – Berkson case. J. Am. Stat. Assoc. **83**(403), 739–743 (1988). MR0963801. doi:10.1080/01621459.1988.10478656

[2] Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M.: Measurement Error in Nonlinear Models: A Modern Perspective. CRC Press (2006). MR2243417. doi:10.1201/9781420010138

[3] Küchenhoff, H.: The identification of logistic regression models with errors in the variables. Stat. Pap. **36**(1), 41–47 (1995). MR1334083. doi:10.1007/BF02926017

[4] Shklyar, S.V.: Logistic regression with homoscedastic errors – A Berkson model. Theory Probab. Math. Stat. **85**, 169–180 (2012). MR2933712. doi:10.1090/S0094-9000-2013-00883-7