Biometrika (2010), **97**, ?, *pp*. 1–2

Printed in Great Britain

Supplementary material for "A goodness-of-fit test of logistic regression models for case-control data with measurement errors" by G. Xu and S. Wang

1. Additional results from the simulation and the application

Additional simulation results supplementary to Table 1 in the paper by Xu and Wang are given in Tables 1 and 2. Estimated coefficients in the Framingham heart disease data analysis and their standard errors are given in Table 3.

Table 1. *Estimation accuracy in Simulation Examples 1 and 3.*

			X log-normal				$X \mid Y = 0$ normal			
			$\hat{\beta}_Z$		$\hat{\beta}_X$		$\hat{\beta}_Z$		$\hat{\beta}_X$	
$\theta = 0$	(n_1, n_2)	Method	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
	$(100, 200)$	M_n	0·06	0·25	-0·07	0·39	0·00	0·14	-0·03	0·19
		D_n	-0·24	0·30	0·47	0·49	0·00	0·14	0·18	0·22
		L_n	-0·24	0·30	0·47	0·49	0·00	0·14	0·18	0·22
	$(200, 100)$	M_n	0·03	0·26	-0·10	0·50	0·00	0·14	-0·03	0·19
		D_n	-0·25	0·29	0·51	0·52	0·00	0·14	0·18	0·23
		L_n	-0·25	0·29	0·51	0·52	0·00	0·14	0·18	0·23

M_n is the test statistic presented in this paper, D_n is the test statistic proposed by Zhang (2001) and L_n is the test statistic proposed by Lin et al. (2002). RMSE stands for root mean square error.

Table 2. *Estimation accuracy in Simulation Example 2.*

			$\hat{\beta}'_Z$		$\hat{\beta}'_X$	
	(n_0, n_1)	Method	Bias	RMSE	Bias	RMSE
$\theta = 0$	$(100, 200)$	D_n	0·003	0·17	-0·01	0·19
	$(200, 100)$	D_n	0·010	0·17	-0·01	0·21

RMSE stands for root mean square error. The true values of coefficients $\beta'_Z = 0{\cdot}4$ and $\beta'_X = -0{\cdot}9$.

Table 3. *Estimated coefficients in the Framingham heart disease data analysis*

	Method	Age	Smoke	Log-Chol	Log-SBP
Estimates	M_n	0·052	0·53	4·02	3·07
Estimates	D_n	0·056	0·53	2·79	2·29
SE	M_n	0·018	0·37	1·19	0·81
SE	D_n	0·017	0·35	0·78	0·58

M_n: our method. D_n: method of Zhang (2001). SE: standard error.

2. Additional proofs

Claim 1: Under error structure (5), model (1) leads to model (6).

Proof. If model (1) is true, then

$$\begin{aligned}\text{pr}(Y = y \mid Z, X, \Delta) &\propto \text{pr}(Y = y \mid Z, X) f(\Delta \mid Y = y, X) \\ &\propto \exp\{y(\beta_0^* + Z^{\mathrm{T}}\beta_Z + \Delta^{\mathrm{T}}\beta_X - \beta_X^{\mathrm{T}}\Sigma_U\beta_X/2)\},\end{aligned}$$

which does not depend on X, where $f(\Delta \mid Y = y, X)$ is the density function of $\Delta \mid Y = y, X$. Similar definitions are used below. Hence

$$\frac{\text{pr}(Y = 1 \mid Z, \Delta)}{\text{pr}(Y = 0 \mid Z, \Delta)} = \exp\{\beta_0^* + Z^{\mathrm{T}}\beta_Z + \Delta^{\mathrm{T}}\beta_X - \beta_X^{\mathrm{T}}\Sigma_U\beta_X/2\},$$

which further implies that

$$\frac{g_1(Z, \Delta)}{g_0(Z, \Delta)} = \exp\{\beta_0 + Z^{\mathrm{T}}\beta_Z + \Delta^{\mathrm{T}}\beta_X - \beta_X^{\mathrm{T}}\Sigma_U\beta_X/2\}.$$

Therefore, model (1) leads to model (6) under error structure (5). □

Proof of Lemma 2. If model (8) is true, then

$$\begin{aligned}&\text{pr}(Y = y \mid Z, X, \Delta) \\ &\quad = \{r^*(Z, X, \eta)\}^y \frac{\exp\{y(\beta_0^* + Z^{\mathrm{T}}\beta_Z + \Delta^{\mathrm{T}}\beta_X - \beta_X^{\mathrm{T}}\Sigma_U\beta_X/2)\}}{f(\Delta \mid Z, X)\{1 + r^*(Z, X, \eta)\exp(\beta_0^* + Z^{\mathrm{T}}\beta_Z + X^{\mathrm{T}}\beta_X)\}}.\end{aligned}$$

Let $r_1^*(Z, X, \Delta, \eta) = f(\Delta \mid Z, X)\{1 + r^*(Z, X, \eta)\exp(\beta_0^* + Z^{\mathrm{T}}\beta_Z + X^{\mathrm{T}}\beta_X)\}$ and $r_2^*(Z, X, \Delta, \eta) = r^*(Z, X, \eta)/r_1^*(Z, X, \Delta, \eta)$. Then

$$\text{pr}(Y = 1 \mid Z, X, \Delta) = r_2^*(Z, X, \Delta, \eta)\exp(\beta_0^* + Z^{\mathrm{T}}\beta_Z + \Delta^{\mathrm{T}}\beta_X - \beta_X^{\mathrm{T}}\Sigma_U\beta_X/2)$$

and

$$\text{pr}(Y = 0 \mid Z, X, \Delta) = 1/r_1^*(Z, X, \Delta, \eta).$$

Thus,

$$\frac{\text{pr}(Y = 1 \mid Z, \Delta)}{\text{pr}(Y = 0 \mid Z, \Delta)} = r(Z, \Delta, \eta)\exp(\beta_0^* + Z^{\mathrm{T}}\beta_Z + \Delta^{\mathrm{T}}\beta_X - \beta_X^{\mathrm{T}}\Sigma_U\beta_X/2),$$

where $r(Z, \Delta, \eta) = \int_{\mathcal{X}} r_2^*(Z, x, \Delta, \eta) f_X(x)\,dx / \int_{\mathcal{X}} \{r_1^*(Z, x, \Delta, \eta)\}^{-1} f_X(x)\,dx$. Since for all (Z, X), $r^*(Z, X, \eta_0) = 1$, one has $r(Z, \Delta, \eta_0) = 1$ for all (Z, Δ). Straightforward algebra shows that

$$\frac{g_1(Z, \Delta)}{g_0(Z, \Delta)} = r(Z, \Delta, \eta)\exp(\beta_0 + Z^{\mathrm{T}}\beta_Z + \Delta^{\mathrm{T}}\beta_X - \beta_X^{\mathrm{T}}\Sigma_U\beta_X/2),$$

which completes the proof. □