Regression Calibration and Tweedie's Formula

Weixing Song

Department of Statistics
Kansas State University

August 17, 2016

Outline

1. Motivation

2. Laplace Measurement Error
   - Parametric Inference with Laplace Measurement Error
   - Hong-Tamer's Estimator
   - Comparison Study

3. Parametric Inference with Normal Measurement Errors
   - Tweedie's Formula

Outline

Motivation

Suppose

- X: The true variable which cannot be observed;
- Z: The surrogate of X;
- Y: Response variable;

  ... Sometimes,
- W: Variables measured without errors;

Regression Calibration

- Estimate the regression of X on Z, $m_X(Z, \gamma)$, depending on parameters γ, which are estimated by $\hat{\gamma}$.
- Replace the unobserved X by its estimate $m_X(Z, \hat{\gamma})$, and then run a standard analysis to obtain parameter estimates.
- Adjust the resulting standard errors to account for the estimation of γ, using either the bootstrap or sandwich method.

For example, suppose that the mean of Y given X can be modeled by

$$E(Y|X) = m_Y(X; \theta)$$

for some unknown parameter θ. Regression calibration is dealing with the following approximate model:

$$E(Y|Z) \approx m_Y(m_X(Z, \gamma), \theta)$$

An extensive discussion on the regression calibration technique can be found in Raymond J. Carroll et al. (2006).

Question?

Note that

$$E(Y|Z) = E[E(Y|X,Z)|Z] = E[E(Y|X)|Z] = E[m_Y(X;\theta)|Z].$$

Instead of "moving the conditional expectation inside m_Y", why not calculate or approximate $E[m_Y(X;\theta)|Z]$ directly?

If this cannot be done "nicely" in the general cases, for example, when $Z = X + U$, and U has a known distribution, can we make it for some special cases, such as U has a normal distribution or a Laplace distribution?

Outline

Multivariate Laplace Distribution:

Multivariate generalization of the Laplace distribution have been considered by many authors. See McGraw and Wagner (1968), Johnson and Kotz (1972), Johnson (1987), Anderson (1992), Osiewalski and Steel (1993), Marshall and Olkin (1993), Kotz, Kozubowski and Podgorski (2001), among others.

However, the term Multivariate Laplace Distribution is still somewhat ambiguous.

A commonly used definition of Laplace distribution is given below.

Definition

A random vector X in $\mathbb{R}^k$ is said to have a multivariate laplace distribution if its characteristic function is given by

$$\phi(t) = \frac{e^{i\mu t}}{1 + t'\Sigma t/2},$$

where $\mu \in \mathbb{R}^k$ and Σ is a $k \times k$ nonnegative definite matrix. Denote $X \sim ML_k(\mu, \Sigma)$.

The Density Function of $ML_k(\mu, \Sigma)$

Suppose $X \sim ML_k(\mu, \Sigma)$.

Denote $Q(x; \mu, \Sigma) = (x - \mu)' \Sigma^{-1} (x - \mu)$.

Density Function of $ML_k(\mu, \Sigma)$

$$f_X(x) = \frac{2}{(2\pi)^{k/2} |\Sigma|^{1/2}} \left[\frac{Q(x; \mu, \Sigma)}{2} \right]^{\frac{1}{2}(1 - \frac{k}{2})} K_{k/2-1}\left(\sqrt{2Q(x; \mu, \Sigma)} \right),$$

where $x \in \mathbb{R}^k$, $K_v(x)$ is the modified Bessel function of the 2nd kind with order v.

Note:

- The modified Bessel function of the second kind.

$$K_v(x) = \frac{\Gamma(v + 1/2)(2x)^v}{\sqrt{\pi}} \int_0^\infty \frac{\cos(t)}{(t^2 + x^2)^{v+1/2}} \, dt.$$

- Let $V \sim$Exp(1), $Z \sim N_k(0, I)$. Then

$$U = \sqrt{V} \Sigma^{1/2} Z \sim ML_k(0, \Sigma).$$

Parametric Inference with Laplace Measurement Error

Suppose that the parameter of interest θ is determined by the following moment condition

$$Em(X;\theta) = 0,$$

where X is a k-dimensional vector, θ is a p-dimensional unknown parameter, and m is a m-dimensional real function.

Sometimes X cannot be observed directly, instead, one can observe

$$Z = X + U,$$

where U is called the measurement error.

Hong-Tamer's Estimator

Suppose that the components of U are independent, and each one follows a Laplace distribution with mean 0.

Denote

$$\bar{m}(Z;\theta,\sigma) = m(Z;\theta) + \sum_{l=1}^{k}\left(-\frac{1}{2}\right)^{l}\sum_{j_1<\cdots<j_l}\sigma_{j_1}^2\cdots\sigma_{j_l}^2\left(\frac{\partial^{2l}m(Z;\theta)}{\partial Z_{j_l}^2\cdots\partial Z_{j_l}^2}\right),$$

Hong and Tamer (2003) showed that, under some smooth conditions on m,

$$E\bar{m}(Z;\theta,\sigma) = Em(X;\theta).$$

Hong and Tamer (2003) proposed the following modified moment estimators,

$$(\hat{\theta},\hat{\sigma}) = \text{argmin}_{\theta,\sigma}\left(\sum_{i=1}^{n}\bar{m}(Z_i;\theta,\sigma)\right)' W_n\left(\sum_{i=1}^{n}\bar{m}(Z_i;\theta,\sigma)\right).$$

Asymptotic normality of the above estimators is derived.

Some limitations in Hong and Tamer (2003)'s estimation procedure

- This method only applies to the cases where the components of U are independent;
- $Em(X;\theta) = Em(Z;\theta,\sigma)$ is an equality of unconditional expectation.

Example (Guo and Li, 2002): Consider the Possion regression model

$$P(Y = y|X) = \frac{\exp(yX\theta)}{y!}\exp(-\exp(X\theta))$$

with measurement error $Z = X + U$. If $E\exp(X\theta)$ is known or can be well estimated, then one can estimate θ using the maximizer of

$$\mathcal{L}_{0n}(\theta) = \frac{1}{n}\sum_{i=1}^{n}[Y_i Z_i\theta - \log Y_i!] - E\exp(X\theta).$$

Applying Hong and Tamer (2004)'s formula,

$$E\exp(X\theta) = \left[1 - \frac{1}{2}\sigma^2\theta^2\right] E\exp(Z\theta).$$

If further assume that σ^2 is known. One can estimate θ using the maximizer of

$$\mathcal{L}_{1n}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_iZ_i\theta - \log Y_i! - \left(1 - \frac{1}{2}\sigma^2\theta^2\right)\exp(Z_i\theta)\right].$$

However, with very large probability, $\mathcal{L}_{1n}(\theta) \to +\infty$ as $|\theta| \to \infty$.

First Improvement

A Relation between Densities of X and Z (Multivariate Case)

Let f, g be the density functions of X and Z, respectively. Assume that $U \sim ML_l(0, \Sigma)$.

Lemma

Assume the characteristic function of Z is square integrable. Then

$$f(x) = g(x) - \frac{1}{2}\sum_{j,l=1}^{k} \sigma_{jl} \frac{\partial^2 g(x)}{\partial x_j \partial x_l},$$

where σ_{jl} is the (j, l)-th element of Σ, and x_j is the j-th element of x.

Based on the above lemma, we have

Theorem

Assume that $m(x, \theta)$ and $g(z)$ satisfy

(C1). For any $\theta \in \Theta$, $m(x, \theta)$ is twice differentiable w.r.t. x; and as $\|x\| \to \infty$,

$$m(x, \theta) g'(x) \to 0, \quad m'(x, \theta) g(x) \to 0.$$

(C2). For any $\theta \in \Theta$,

$$E\|m(Z, \theta)\| < \infty, \quad E\left\|\frac{\partial^2 m(Z, \theta)}{\partial Z_j \partial Z_l}\right\| < \infty,$$

Then we have

$$Em(X, \theta) = Em(Z, \theta) - \frac{1}{2} \sum_{j,l=1}^{k} \sigma_{jl} E \frac{\partial^2 m(Z, \theta)}{\partial Z_j \partial Z_l}.$$

Denote

$$\widetilde{m}(Z;\theta,\Sigma) = m(Z,\theta) - \frac{1}{2}\sum_{j,l=1}^{k} \sigma_{jl} \frac{\partial^2 m(Z,\theta)}{\partial Z_j \partial Z_l}.$$

For any positive definite matrix, depending only on the data, define

$$(\hat{\theta},\hat{\Sigma}) = \text{argmin}_{\theta,\Sigma} \left(\sum_{i=1}^{n} \widetilde{m}(Z_i;\theta,\Sigma) \right)' W_n \left(\sum_{i=1}^{n} \widetilde{m}(Z_i;\theta,\Sigma) \right). \quad (1)$$

We further assume that

(C3). $E\widetilde{m}(Z;\theta,\Sigma) = 0$ if and only if $\theta = \theta_0$, $\Sigma = \Sigma_0 > 0$, where θ_0, Σ_0 are the true values.

(C4). $W_n \to W$ in probability;

(C5). Denote $\sigma = (\sigma_{jl}, j \geq l)$, $\alpha = (\theta', \sigma')'$. $E\partial\widetilde{m}(Z;\theta,\Sigma)/\partial\alpha$ exists, and has the full rank; In a neighborhood of α_0, $\partial\widetilde{m}(Z;\theta,\Sigma)/\partial\alpha$ is Lipschitz continuous. $E\|\widetilde{m}(Z;\theta_0,\Sigma_0)\|^2 < \infty$.

Theorem

Assume that (C1)-(C5) hold. Then $(\hat{\theta}, \hat{\sigma}) \to (\theta_0, \sigma_0)$ in probability, also

$$\sqrt{n}\begin{pmatrix}\hat{\theta}-\theta_0\\ \hat{\sigma}-\sigma_0\end{pmatrix} \Longrightarrow N\left(0, (A'WA)^{-1}(A'W\Omega WA)(A'WA)^{-1}\right),$$

where

$$A = \left[\frac{\partial m(Z;\theta)}{\partial\theta} - \frac{1}{2}\sum_{j,l=1}^{k}\sigma_{jl}\frac{\partial m^3(Z;\theta)}{\partial Z_j \partial Z_l \partial\theta}, \quad \left(-\frac{1}{2^{\delta(j,l)}}\frac{\partial^2 m(Z;\theta)}{\partial Z_j \partial Z_l}, j \geq l\right)\right],$$

$\Omega = E\widetilde{m}(Z;\theta,\Sigma)\widetilde{m}'(Z;\theta,\Sigma)$; if $j = l$, $\delta(j,l) = 1$, otherwise, $\delta(j,l) = 0$.

From the consistency of $\hat{\theta}, \hat{\Sigma}$, we also have

Theorem

Assume that (C1)-(C5) hold. Let $\hat{W}_n = n(\sum_{i=1}^n \widetilde{m}(Z_i; \hat{\theta}, \hat{\Sigma})\widetilde{m}'(Z_i; \hat{\theta}, \hat{\Sigma}))^{-1}$, and

$$(\widetilde{\theta}, \widetilde{\Sigma}) = \text{argmin}_{\theta,\sigma} \left(\sum_{i=1}^n \widetilde{m}(Z_i; \theta, \sigma) \right)' \hat{W}_n \left(\sum_{i=1}^n \widetilde{m}(Z_i; \theta, \sigma) \right).$$

Then, $\hat{W}_n \to \Omega^{-1}$ in probability, $(\widetilde{\theta}, \widetilde{\Sigma}) \to (\theta_0, \Sigma_0)$ in probability. Furthermore,

$$\sqrt{n} \begin{pmatrix} \widetilde{\theta} - \theta_0 \\ \widetilde{\sigma} - \sigma_0 \end{pmatrix} \Longrightarrow N\left(0, (A'\Omega A)^{-1}\right),$$

where $\widetilde{\sigma} = (\widetilde{\sigma}_{jl}, j \geq l)$, $\widetilde{\sigma}_{jl}$ is the (j, l)-th element of $\tilde{\Sigma}$.

Second Improvement: Generalized Regression Calibration

Tweedie-type Formula for Laplace Distribution

For the time being, θ will be suppressed from $m(x, \theta)$.

Denote

- g: the density function of Z;
- f_U: the density function of U.

If $U \sim ML_k(0, \Sigma)$, we have

Tweedie-type Formula for Laplace Distribution

Under some regularity conditions,

$$E[m(X)|Z] = \frac{1}{g(Z)} \left[\int m(x) f_U(Z-x) g(x)\, dx - \frac{1}{2} \sum_{j,l=1}^{n} \sigma_{jl} \int m(x) f_U(Z-x) \frac{\partial^2 g(x)}{\partial x_j \partial x_l}\, dx \right].$$

In particular, for $k = 1$, we have

Corollary

Suppose m and the density function g of Z are twice differentiable. For $w(x) = m(x)g(x)$, $m'(x)g(x)$, $m(x)g'(x)$, we further assume that

$$\lim_{x \to \pm\infty} w(x)\exp(-|x|) \to 0.$$

Then,

$$E[m(X)|Z] = m(Z) + l(Z),$$

where

$$l(z) = \frac{e^{z/b}}{g(z)} \int_z^{\infty} \left[m'(x) - \frac{bm''(x)}{2} \right] g(x) e^{-x/b} \, dx - \frac{e^{-z/b}}{g(z)} \int_{-\infty}^{z} \left[m'(x) + \frac{bm''(x)}{2} \right] g(x) e^{x/b} \, dx,$$

and $b = \sigma/\sqrt{2}$.

For the sake of brevity, we only consider the case of $k = 1$.

Let $Z_1, Z_2, \ldots, Z_n$ be a sample from Z, and let

$$\hat{g}_h(z) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{Z_i - z}{h}\right).$$

be the kernel density function of Z. Denote

$$\mu_l(x;\theta,\sigma) = \left[m'(x;\theta) + (-1)^l \frac{\sigma}{2\sqrt{2}} m''(x;\theta)\right] \exp\left((-1)^l \frac{x\sqrt{2}}{\sigma}\right),$$

then for any fixed θ, σ, $E(m(X,\theta)|Z)$ can be estimated by

$$\widehat{E}(m(X,\theta)|Z) = m(Z;\theta) + \frac{e^{Z\sqrt{2}/\sigma}}{\hat{g}_h(Z)} \int_Z^\infty \mu_1(x;\theta,\sigma)\hat{g}_h(x)\,dx - \frac{e^{-Z\sqrt{2}/\sigma}}{\hat{g}_h(Z)} \int_{-\infty}^{Z} \mu_2(x;\theta,\sigma)\hat{g}_h(x)\,dx.$$

Let $H(Z;\theta,\sigma) = E(m(X,\theta)|Z)$. Thus for fixed θ,σ, we can estimate $H(Z;\theta,\sigma)$ using

$$\hat{H}(Z;\theta,\sigma) = \hat{E}(m(X,\theta)|Z).$$

For any positive definite matrix W_n, depending only on the data, we estimate θ and σ by

$$(\hat{\theta},\hat{\sigma}) = \text{argmin}_{\theta,\sigma} \left(\sum_{i=1}^{n} \hat{H}(Z_i;\theta,\sigma) \right)' W_n \left(\sum_{i=1}^{n} \hat{H}(Z_i;\theta,\sigma) \right).$$

<u>Work to do:</u> Asymptotic theory.

Comparison Study

Consider the Possion regression model

$$P(Y = y|X) = \frac{\exp(yX\theta)}{y!} \exp(-\exp(X\theta))$$

with measurement error $Z = X + U$. If $E\exp(X\theta)$ is known or can be well estimated, Guo and Li (2002) suggests to estimate θ using the maximizer of

$$\mathcal{L}_{0n}(\theta) = \frac{1}{n}\sum_{i=1}^{n}[Y_i Z_i \theta - \log Y_i!] - E\exp(X\theta).$$

Take $\theta = 1$, $X \sim N(0, 1)$, and $U \sim$Laplace(0,σ^2).

The density function of X is known

σ^2	n	Guo and Li	Regression Calibration
0.1	100	0.0318(-0.0290)	0.0064(-0.0086)
	200	0.0172(-0.0188)	0.0031(-0.0104)
	300	0.0125(-0.0117)	0.0022(-0.0090)
	500	0.0059(-0.0061)	0.0013(-0.0033)
0.5	100	0.0354(-0.0354)	0.0172(-0.0264)
	200	0.0189(-0.0099)	0.0085(-0.0083)
	300	0.0128(-0.0111)	0.0063(-0.0092)
	500	0.0085(-0.0088)	0.0038(-0.0041)
0.9	100	0.0489(-0.0423)	0.0328(-0.0376)
	200	0.0211(-0.0265)	0.0138(-0.0236)
	300	0.0176(-0.0150)	0.0117(-0.0143)
	500	0.0083(-0.0125)	0.0055(-0.0132)

Table: MSE(Bias) Comparison

The density function of X is unknown

Hong-Tamer Technique

σ^2	n	MSE(Bias)
0.1	100	87.7185(5.6400)
	200	88.1186(5.4400)
	300	86.9986(5.9999)
	500	85.6386(6.6799)
0.5	100	94.5986(2.2000)
	200	95.0786(1.9600)
	300	96.3586(1.3200)
	500	97.0786(0.9600)
0.9	100	100.9985(-1.0000)
	200	100.7585(-0.8800)
	300	101.1585(-1.0800)
	500	100.4385(-0.7200)

Regression Calibration

$b = \sigma/\sqrt{2}$, $h = an^{-1/5}$.

n	b	a	MSE(Bias)
100	0.2	0.2	0.0062(-0.0358)
		0.5	0.0078(-0.0240)
		0.8	0.0076(-0.0235)
	0.5	0.2	0.0529(-0.1829)
		0.5	0.0536(-0.1104)
		0.8	0.0497(-0.1399)
	0.8	0.2	0.1881(-0.3549)
		0.5	0.2351(-0.2281)
		0.8	0.3000(-0.0787)
200	0.2	0.2	0.0033(-0.0210)
		0.5	0.0038(-0.0173)
		0.8	0.0034(-0.0245)
	0.5	0.2	0.0405(-0.1724)
		0.5	0.0354(-0.1341)
		0.8	0.0309(-0.1265)
	0.8	0.2	0.1566(-0.3573)
		0.5	0.2486(-0.1476)
		0.8	0.2368(-0.1268)

n	b	a	MSE(Bias)
300	0.2	0.2	0.0024(-0.0232)
		0.5	0.0027(-0.0229)
		0.8	0.0026(-0.0222)
	0.5	0.2	0.0341(-0.1525)
		0.5	0.0226(-0.1103)
		0.8	0.0220(-0.1003)
	0.8	0.2	0.1351(-0.3350)
		0.5	0.2106(-0.1728)
		0.8	0.2640(-0.0523)
500	0.2	0.2	0.0018(-0.0213)
		0.5	0.0015(-0.0169)
		0.8	0.0018(-0.0210)
	0.5	0.2	0.0337(-0.1633)
		0.5	0.0208(-0.1059)
		0.8	0.0172(-0.0876)
	0.8	0.2	0.1385(-0.3410)
		0.5	0.1624(-0.1977)
		0.8	0.2250(-0.0674)

Outline

Tweedie's Formula

As disclosed in Efron (2011), the Tweedie's formula is named after Maurice Kenneth Tweedie and it was first discussed in Herbert Robbins (1956). Due to its strong Bayesian flavor, Efron (2011) accolade the Tweedie's formula as an "extraordinary Bayesian estimation formula", and a selection bias application of this formula to genomics data is also discussed.

Tweedie's Formula

Suppose $Z = X + U$, and $U \sim N(0, \sigma^2)$, then

$$E(X|Z) = Z + \sigma^2 \frac{g'(Z)}{g(Z)}.$$

Generalized Tweedie's Formulae

Result 1:

$$\begin{aligned} E(X^2|Z) &= Z^2 + 2\sigma^2 Z \frac{g'(Z)}{g(Z)} + \sigma^4 \frac{g''(Z)}{g(Z)} + \sigma^2; \\ E(X^3|Z) &= Z^3 + 3\sigma^2 Z^2 \frac{g'(Z)}{g(Z)} + 3\sigma^2 Z + 3\sigma^4 Z \frac{g''(Z)}{g(Z)} + \sigma^6 \frac{g'''(Z)}{g(Z)} + 3\sigma^4 \frac{g'(Z)}{g(Z)}; \\ E(X^k|Z) &= ? \quad k \geq 4. \end{aligned}$$

Result 2:

$$E(\exp(\beta X)|Z) = \frac{g(Z+\beta\sigma^2)}{g(Z)} \exp\left(Z\beta + \frac{\beta^2\sigma^2}{2}\right).$$

The above results can be readily applied to some nonlinear regression models.

For example $Y = m(X) + \varepsilon$, when $m(x) =$polynomial, exponential.

Can we apply the above results to parametric/nonparametric setup?

For example,

- Logistic regression. Instead of replacing $\exp(X\beta)$ with $\exp(E(X|Z)\beta)$, try $E(\exp(X\beta)|Z)$;
- if X is observable, the local linear/polynomial regression estimator of $m(x_0)$ can be obtained by minimizing

$$\sum_{i=1}^{n}[Y_i - \alpha_0 - \alpha_1(X_i - x_0) - \alpha_2(X_i - x_0)^2 - \cdots - \alpha_p(X_i - x_0)^p]^2 K_h(X_i - x_0).$$

How about replacing $(X_i - x_0)^j$ with $E[(X_i - x_0)^j|Z_i]$, and replacing $K_h(X_i - x_0)$ with $K_h(E(X_i|Z_i) - x_0)$?

Thank You!