Solutions to Voleon Problem

Yi Gao*
(Dated: December 28, 2016)

I. PATTERN OF DATA

Since the model is linear of the form

$$y_i = ax_i + b + e_i, \tag{1}$$

in which b is the intercept, and a is the linear coefficient. Meanwhile, the noise has property

$$\mathbb{E}\left(e_i \mid x_i\right) = 0, \tag{2}$$

and also the conditional variance of e_i is depending on x_i, i.e.,

$$\mathbb{V}\left(e_i \mid x_i\right) = \sigma^2\left(x_i\right). \tag{3}$$

Therefore, the linear model is heteroskedastic, and the ordinary linear least square (OLS) is no longer the best linear unbiased estimator (BLUE).

To see the pattern of the noise terms, we first use OLS to fit the data in each data set. We then use the residuals of OLS, $\widehat{e}_i$, as an approximation of noise terms, and see the relation between noise and predictors x_i's. We fit each data set using OLS, and regress the square of residuals on to the terms $|x_i|$ and x_i^2,

$$\widehat{e}_i^2 = \alpha + \beta\left|x_i\right| + \beta x_i^2. \tag{4}$$

The results are summarized in Table I.

	Value	t-stat	p-value
α	-88.6353	-1.469	0.143
β	54.4057	3.024	0.003
γ	-1.1313	-1.240	0.216

TABLE I: Regression results of OLS residuals on $|x_i|$ and x_i^2.

According to the above regression result, it can be concluded that the model is indeed heteroskedastic, i.e., the noise term is depending on x_i. Moreover, the regression result strongly suggests the relation as $\mathbb{E}\left(e_i^2 \mid x_i\right) \propto |x_i|$. We now plot $\widehat{e}_i / \left|x_i\right|^{1/2}$ as a function of x_i's in Fig. 1. We can reasonably hypothesize that $e_i / \left|x_i\right|^{1/2}$ is white noise. Note that here e_i is the true noise, instead of OLS residuals.

II. HETEROSKEDASTIC LINEAR MODEL

According to the analysis of data pattern, the noise in the linear model we propose is

$$e_i \mid x_i \sim \mathcal{N}\left(0, c^2\left|x_i\right|\right), \tag{5}$$

i.e., the conditional distribution of noise given x_i is normal, and the variance is proportional to $|x_i|$. We can write the conditional pdf of e_i as

$$f\left(e_i \mid x_i\right) = \frac{1}{\sqrt{2\pi c^2\left(\left|x_i\right| + \eta\right)}} \exp\left(-\frac{e_i^2}{2c^2\left(\left|x_i\right| + \eta\right)}\right). \tag{6}$$

*Electronic address: yi_gao@mfe.berkeley.edu

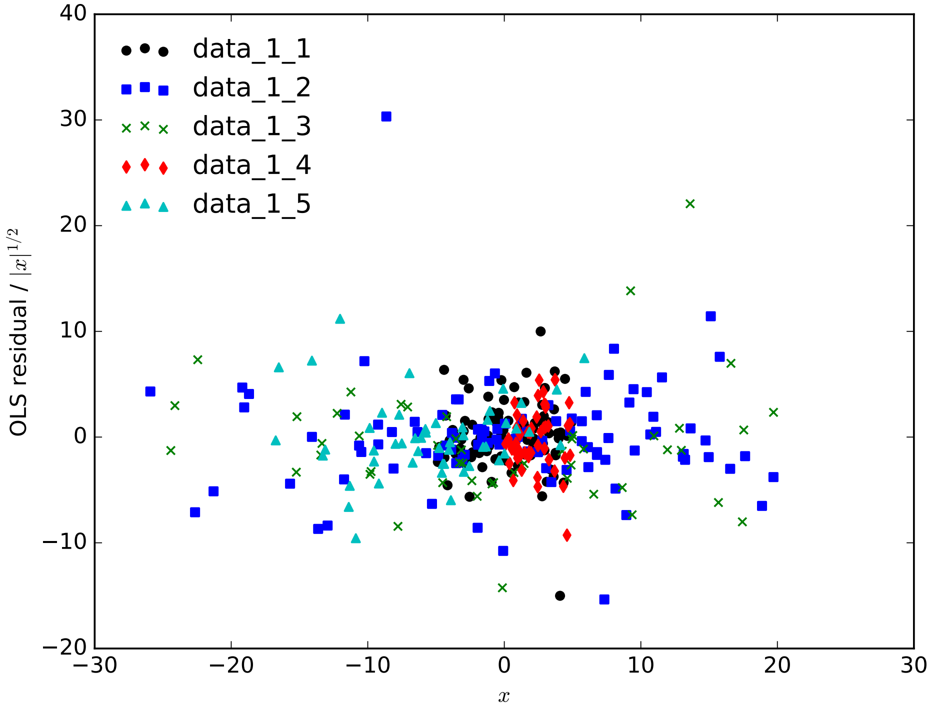

FIG. 1: $\widehat{e}_i / |x_i|^{1/2}$as a function of predictors x_i's.

Here we used $\eta \to 0^+$ to avoid divengence as $x_i \to 0$. The log-likelihood function of each data set can be written as

$$\widehat{l}_n (a, b, c \mid x) = \frac{1}{n} \sum_{i=1}^{n} \ln f (y_i - a x_i - b \mid x_i), \tag{7}$$

and in machine learning framework, we can formulate loss function J as

$$J (a, b, c) = -\widehat{l}_n (a, b, c \mid x) = \frac{1}{2} \ln \left(2\pi c^2\right) + \frac{1}{2n} \sum_{i=1}^{n} \ln (|x_i| + \eta) + \frac{1}{2nc^2} \sum_{i=1}^{n} \frac{(y_i - a x_i - b)^2}{|x_i| + \eta}. \tag{8}$$

The point estimate of parameters can be obtained by minimizing loss function J. The minimum can be obtained with condition

$$\begin{cases} \frac{\partial}{\partial a} J &= 0, \\ \frac{\partial}{\partial b} J &= 0, \\ \frac{\partial}{\partial c} J &= 0. \end{cases} \tag{9}$$

For a and b, we can solve as

$$\begin{aligned} \sum_{i=1}^{n} \frac{(y_i - a x_i - b)\, x_i}{|x_i| + \eta} &= 0, \\ \sum_{i=1}^{n} \frac{y_i - a x_i - b}{|x_i| + \eta} &= 0. \end{aligned}$$

In matrix form, a and b are obtained as

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} \frac{x_i^2}{|x_i|+\eta} & \sum_{i=1}^{n} \frac{x_i}{|x_i|+\eta} \\ \sum_{i=1}^{n} \frac{x_i}{|x_i|+\eta} & \sum_{i=1}^{n} \frac{1}{|x_i|+\eta} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{n} \frac{x_i y_i}{|x_i|+\eta} \\ \sum_{i=1}^{n} \frac{y_i}{|x_i|+\eta} \end{pmatrix}. \tag{10}$$

When a and b are obtained, it is straightforward that

$$c = \sqrt{\sum_{i=1}^{n} \frac{(y_i - a x_i - b)^2}{|x_i| + \eta}}. \tag{11}$$

Fitting each data set, the point estimate of a and b are summarized