*Journal of the Korean Data & Information Science Society*
*2017*, **28**(5), 1191–1204

http://dx.doi.org/10.7465/jkdi.2017.28.5.1191
한국데이터정보과학회지

# A study on robust regression estimators in heteroscedastic error models[†]

Nayeong Son[1] · Mijeong Kim[2]

[12]Department of Statistics, Ewha Womans University

Received 1 September 2017, revised 15 September 2017, accepted 18 September 2017

**Abstract**

Weighted least squares (WLS) estimation is often easily used for the data with heteroscedastic errors because it is intuitive and computationally inexpensive. However, WLS estimator is less robust to a few outliers and sometimes it may be inefficient. In order to overcome robustness problems, Box-Cox transformation, Huber's $M$ estimation, bisquare estimation, and Yohai's MM estimation have been proposed. Also, more efficient estimations than WLS have been suggested such as Bayesian methods (Cepeda and Achcar, 2009) and semiparametric methods (Kim and Ma, 2012) in heteroscedastic error models. Recently, Çelik (2015) proposed the weight methods applicable to the heteroscedasticity patterns including butterfly-distributed residuals and megaphone-shaped residuals. In this paper, we review heteroscedastic regression estimators related to robust or efficient estimation and describe their properties. Also, we analyze cost data of U.S. Electricity Producers in 1955 using the methods discussed in the paper.

*Keywords*: Bayesian heteroscedastic regression model, heteroscedastic regression, robust estimators, semiparametric efficient estimators, weighting absolute centered external variable.

## 1. Introduction

We consider a regression model as follows.

$$Y = \boldsymbol{X}\boldsymbol{\beta} + \epsilon, \tag{1.1}$$

where $Y$ is a one-dimensional response variable, $\boldsymbol{X}$ is a $n \times p$ design matrix, $\boldsymbol{\beta}$ is a $p$-dimensional parameter and $n$ is the number of observations. In the above regression model, ordinary least squares (OLS) estimation is often easily used. OLS estimation assumes that $\epsilon_i$ has mean zero $E(\epsilon_i) = 0$ and equal variance $\text{Var}(\epsilon_i) = \sigma^2$ for $i = 1, \cdots, n$. The model based

[†] This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean Government (NRF-2017R1C1B5015186).
[1] Graduate student, Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul, Republic of Korea. E-mail: sonnayeong@ewhain.net
[2] Corresponding author: Assistant professor, Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul, Republic of Korea. E-mail: m.kim@ewha.ac.kr

on equal variances is called a homoscedastic error model. If the model error $\epsilon$ satisfies the normal assumption and has the equal variance given $\boldsymbol{X}$, OLS estimator is efficient in the class of unbiased estimators and identical to maximum likelihood estimator (MLE). In this sense, we call OLS estimator best linear unbiased estimator (BLUE) in linear regression models. However, we often encounter the situations that $\epsilon$ does not satisfy normal assumption or $\epsilon$ increases as $\boldsymbol{X}$ increases. The model based on unequal variances given covariates, we call a heteroscedastic error model. For the heteroscedasic errors, we often use weighted least squares (WLS) estimation. In WLS estimation, $E(\epsilon_i) = 0$ and $\text{Var}(\epsilon_i) = \sigma/w_i$ are assumed, where $w_i$'s are weights for $i = 1, \cdots, n$. Weighted version of the residual sum of squares (WRSS) is given by

$$\text{WRSS} = \sum_{i=1}^{n} w_i (y_i - \boldsymbol{x}_i^{\text{T}} \boldsymbol{\beta})^2,$$

where $\boldsymbol{x}_i$ is the covariate vector of $i^{th}$ observation. Minimizing the WRSS, we obtain WLS estimators. WLS estimation is intuitive and computationally inexpensive, but WLS estimator is not robust nor efficient for the data with heteroscedastic errors. Sometimes WLS may mislead interpretation.

Robust and efficient estimations have been considered as a major issue in heteroscedastic error models. The very simple approach for robust estimation is to use WLS estimation after deleting some outliers because the WLS estimation performs poorly when a few outliers are detected. Box and Cox (1964) proposed the transformation method of the response variable. Huber's $M$ estimation (Huber, 1973) and Yohai's MM estimation (Yohai, 1987) provide the robust estimators for heteroscedastic error data. Cepeda and Achcar (2009) performs Bayesian approaches to find robust estimators in heteroscedastic error models. Kim and Ma (2012) proposes efficient estimators in semiparametric heteroscedastic error models with the regression mean and variance functions. Recently, Çelik (2015) suggested a robust methods to stabilize heteroscedasticity for butterfly distributed residuals by adjusting weights. In section 2, we briefly describe the traditional robust estimation: Box-Cox method, Huber's $M$ estimation, Bisquare estimation and Yohais MM estimation. And we also review three heteroscedastic error models: and Bayesian estimators (Cepeda and Achcar, 2009), semiparametric efficient (SE) estimator (Kim and Ma, 2012) and weighting absolute centered external variable (WCEV) (Çelik, 2015) estimators. Unless otherwise specified, (1.1) is used as a main model in those methods. In section 3, we analyze cost data of 145 U.S. Electricity Producers in 1955. In section 4, we conclude the paper.

## 2. Methods

### 2.1. Robust estimation

#### 2.1.1. Box-Cox transformation on the response

Box and Cox (1964) proposed the method to transform the response $y$ to stabilize heteroscedastic structure. They make use of $y^\lambda$ or $\log(y)$ instead of $y$ in the regression model and estimate $\lambda$ and $\boldsymbol{\beta}$.

$$w = \begin{cases} \dfrac{y^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0, \\ \log(y) & \text{if } \lambda = 0. \end{cases}$$

We regress the transformed response $w$ on covariates in the following way.

$$w = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta} + \epsilon,$$

where $\boldsymbol{X}$ is a design matrix. Here, the parameter $\boldsymbol{\beta}$ and $\lambda$ are estimated simultaneously by maximizing likelihood function. The Box-Cox transformation can be applied to the data with the various relation between the response and the explanatory variables. However, it is difficult to interpret the original response variable using the covariates when the response is transformed through Box-Cox transformation.

### 2.1.2. Huber's $M$ and bisquare estimation

Huber's $M$ and bisquare estimation are designed to calculate optimal weights which reduce the influence of outliers. Huber's $M$ estimators is obatined from minimizing $\sum_{i=1}^{n} \rho(e_i)$, where

$$\rho(e_i) = \begin{cases} e_i^2/2, & \text{if } |e_i/c| \leq 1, \\ c|e_i| - c^2/2 & \text{if } |e_i/c| > 1. \end{cases}$$

That is, if $e_i$ is greater than $c$, this methods weaken the influence of $e_i$. The value $c$ is called a tuning constant. The Huber's $M$ estimators are less affected from outliers when $c$ is smaller. Let $\psi(x)$ be a first derivative of $\rho$ with respect to $e_i$.

$$\psi(e_i) = \begin{cases} e_i, & \text{if } |e_i/c| \leq 1, \\ c \cdot \mathrm{sign}(e_i), & \text{if } |e_i/c| > 1. \end{cases}$$

From the above, the weights are given below.

$$\begin{cases} 1, & \text{if } |e_i/c| \leq 1, \\ c/|e_i|, & \text{if } |e_i/c| > 1. \end{cases}$$

Bisquare estimation is also known as Tukey's bisquare estimation or biweight estimation and adopt $M$ estimator similar to Huber's methods. Bisquare's objective function $\rho(e_i)$ and its derivative function $\psi(e_i)$ as follows.

$$\begin{aligned} \rho(e_i) &= \begin{cases} (c^2/6)[1 - \{1 - (e_i/c)^2\}^3], & \text{if } |e_i/c| \leq 1, \\ c^2/6, & \text{if } |e_i/c| > 1, \end{cases} \\ \psi(e_i) &= \begin{cases} e_i\{1 - (e_i/c)^2\}^2, & \text{if } |e_i/c| \leq 1, \\ 0, & \text{if } |e_i/c| > 1. \end{cases} \end{aligned}$$

The weights of bisquare estimation are given by

$$\begin{cases} \left\{1 - (e_i/c)^2\right\}^2, & \text{if } |e_i/c| \leq 1, \\ 0, & \text{if } |e_i/c| > 1. \end{cases}$$

Any symmetric and nonnegative functions can be used as $\rho(\cdot)$ if $\rho(0) = 0$ and those functions are monotone on the absolute value of residual. Note that $c = 1.345 \times \sigma$ is frequently used (where $\sigma$ is the standard deviation of the errors) when the error term is normally distributed, which satisfies 95% efficiency (Huber, 1973). R package robustreg provides $M$-estimators using iteratively reweighted least squares (IRLS) such as Huber's $M$ estimator (Huber, 1973) or bisquare estimator.

### 2.1.3. Yohai's MM estimation

Yohai (1987) proposed MM estimation to obtain efficient estimators robust to outliers in regression models. MM estimators take advantages of efficiency for the estimates with high breakdown-point. The MM estimators can be obtained thorough the three stages. In the first stage, we obtain initial regression estimators through $S$-estimation proposed in Rousseeuw and Yohai (1987). Yohai (1987) remarks the initial estimators of the first stage is consistent and robust, but it is not always efficient. In the second stage, $M$ estimators for the errors scale is calculated using the residuals obtained from initial estimation. Note that the scale estimate $s(u)$ is a root of (2.1).

$$(1/n) \sum_{i=1}^{n} \rho(u_i/s) = b, \tag{2.1}$$

where $E_\phi\{\rho(u)\}$ can be used as $b$ and $\phi$ denotes the standard normal distribution. Here, $\rho$ is a continuous monotone increasing bounded function satisfies $\rho(0) = 0$ and $\rho(-u) = \rho(u)$. In the third stage, $M$ estimator $\boldsymbol{\beta}_M$ for the regression are obtained using proper $\psi$ function in the below. Finally $\boldsymbol{\beta}_M$ is called a MM estimator.

$$\sum_{i=1}^{n} \boldsymbol{x}_i \psi\{r_i(\boldsymbol{\beta}_M)/s_n\} = 0,$$

where $s_n = s(r_i)$ and $r_i$'s are residuals obtained from the first stage. Here, $\psi$ is the first derivative of the objective function describe in section 2.1.2. The function $\psi(t)$ is also called a redescending score function. Using iteratively reweighted least squares (IRLS) lead to the solution of the above equation. Yohai (1987) pointed out MM estimators are consistent, robust and efficient and also satisfy asymptotic normality.

## 2.2. Heteroscedastic regression models

### 2.2.1. Bayesian heteroscedasticity model

We consider the regression model as follows.

$$Y = m(\boldsymbol{X}, \boldsymbol{\beta}) + \epsilon,$$

where $Y$ is a response, $\boldsymbol{X}$ is design matrix, $\boldsymbol{\beta}$ is a parameter vector, $\epsilon$ is an error and $m$ is a mean function. The error satisfy the followings.

$$E(\epsilon|\boldsymbol{X}) = 0, \quad E(\epsilon^2|\boldsymbol{X}) = \sigma^2(\boldsymbol{X}, \boldsymbol{\gamma}),$$

where $\boldsymbol{\gamma}$ is a parameter vector. Here, we call $\sigma^2$ variance function.

Using a fully Bayesian approach to regression models with heteroscedastic errors, Cepeda and Gamerman (2000) suggested Bayesian estimation of joint regression for mean and variance functions. Since then, Bayesian methodologies for joint regression has been extended to Beta regression, Gamma regression. Cepeda and Gamerman (2005) and Cepeda and Achcar (2009) described detailed algorithms for the joint regression model with normal error distribution and extended the approach to hierarchical regression model.

We first assign multivariate normal distribution as a prior distribution of ($\boldsymbol{\beta}$,$\boldsymbol{\gamma}$). In many simple Bayesian joint regressions, it is often used to sample $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ iteratively from their conditional distribution. However, using conditional distribution for sampling $\boldsymbol{\gamma}$ may result in inefficient estimators in heteroscedastic model. Instead of sampling $\boldsymbol{\gamma}$ from its conditional distribution $\pi(\boldsymbol{\gamma}|\boldsymbol{\beta}, \text{data})$, Cepeda and Gamerman (2000) proposed a transition kernel as (2.2) and adopted Metropolis-Hastings (M-H) algorithm for $\boldsymbol{\gamma}$ in (2.3). Consequently, we generate $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ using mixed sampling process: sample $\boldsymbol{\beta}$ using Gibbs sampling first and then sample $\boldsymbol{\gamma}$ using M-H algorithm. The sampling methods are described in (2.2) and (2.3).

- Prior for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$

$$\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix} \sim \begin{bmatrix} b \\ g \end{bmatrix} + \begin{bmatrix} B & C^{\mathrm{T}} \\ C^{\mathrm{T}} & G \end{bmatrix}.$$

- Sampling distribution for $\boldsymbol{\beta}$

$$\begin{aligned} \boldsymbol{\beta}, \text{data}|\boldsymbol{\gamma} &\sim N(b^*, B^*), \\ \text{where } b^* &= B^*(B^{-1}b + \boldsymbol{X}^{\mathrm{T}}\Sigma^{-1}Y), \\ B^* &= (B^{-1} + \boldsymbol{X}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{X})^{-1}, \\ \Sigma &= \text{diag}\{\sigma_1^2, \sigma_2^2, ..., \sigma_n^2\}. \end{aligned} \tag{2.2}$$

- Sampling distribution for $\boldsymbol{\gamma}$

$$\begin{aligned} \boldsymbol{\gamma}|\boldsymbol{\beta} &\sim N(g^*, G^*), \\ \text{where } g^* &= G^*(G^{-1}g + \boldsymbol{X}^{\mathrm{T}}\Sigma^{-1}\tilde{Y}), \\ G^* &= (G^{-1} + \boldsymbol{X}^{\mathrm{T}}\Sigma^{-1}\boldsymbol{X})^{-1}, \\ \Sigma &= 2I_n, \\ I_n &= \text{diag}\{1, \cdots, 1\}, \\ \tilde{y}_i &= \eta_i + \frac{(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})^2}{\sigma^2(\boldsymbol{x}_i, \boldsymbol{\gamma})} - 1. \end{aligned} \tag{2.3}$$

In the above, $\tilde{Y} \sim N(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\gamma})$, known as "working observational model" is used instead of Y in sampling $\boldsymbol{\gamma}$. While it considers only the case of independent errors, the extended versions of independence model are implemented in Cepeda-Cuervo and Núnez-Antón (2007). Although Bayesian estimation is computationally more expensive than non-bayesian methods, it is very useful in terms of inferences because highest posterior density (HPD) interval and Bayesian hypothesis test are applicable. Also, the heteroscedastic Bayesian model is transmutable into generalized linear model such as joint mean and variance beta regression model (Cepeda-Cuervo, 2015).

#### 2.2.2. Semiparametric efficient (SE) estimation

Kim and Ma (2012) proposed semiparametric efficient estimation for heteroscedastic error models. The considered models is given by

$$Y = m(\boldsymbol{X}, \boldsymbol{\beta}) + \epsilon,$$

where $m$ is a known mean function. The mean function $m$ could be both linear and nonlinear to $\boldsymbol{X}$. They assumed $E(\epsilon|\boldsymbol{X}) = 0$ and $\text{Var}(\epsilon|\boldsymbol{X}) = \sigma^2(\boldsymbol{X}, \boldsymbol{\gamma})$ where the variance function $\sigma^2$ is a known function up to an unknown parameter $\boldsymbol{\gamma}$. Since the model considered in Kim and Ma (2012) includes parameters of interest $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$ and the nonparametric components $\epsilon$, the model is a semiparametric model. Note that $\epsilon$ is a nonparametric component since any distribution assumption on $\epsilon$ is not imposed. According to Tsiatis (2006), Kim and Ma (2012) constructed the nuisance tangent space $\Lambda$ and its orthogonal complement $\Lambda^{\mathrm{T}}$. Projecting score function for $\boldsymbol{\theta}$ on $\Lambda^{\mathrm{T}}$, the efficient score functions $S_{\text{eff}}(\boldsymbol{X}, \boldsymbol{\theta}) = \left\{S_{\boldsymbol{\beta},\text{eff}}(\boldsymbol{X}, Y)^{\mathrm{T}}, S_{\boldsymbol{\gamma},\text{eff}}(\boldsymbol{X}, Y)^{\mathrm{T}}\right\}^{\mathrm{T}}$ can be made.

$$\begin{aligned} S_{\boldsymbol{\beta},\text{eff}}(\boldsymbol{X}, Y) &= \frac{\partial m(\boldsymbol{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \left\{ \frac{\epsilon}{\sigma^2(\boldsymbol{X}, \boldsymbol{\gamma})} - \frac{E(\epsilon^3|\boldsymbol{X})C}{\sigma^2(\boldsymbol{X}, \boldsymbol{\gamma})E(C^2|\boldsymbol{X})} \right\}, \quad (2.4) \\ S_{\boldsymbol{\gamma},\text{eff}}(\boldsymbol{X}, Y) &= \frac{C}{E(C^2|\boldsymbol{X})} \frac{\partial \sigma^2(\boldsymbol{X}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}, \end{aligned}$$

where $C = \epsilon^2 - \sigma^2(\boldsymbol{X}, \boldsymbol{\gamma}) - E\{\epsilon^3 - \epsilon\sigma^2(\boldsymbol{X}, \boldsymbol{\gamma})|\boldsymbol{X}\}\epsilon/\sigma^2(\boldsymbol{X}, \boldsymbol{\gamma})$. By solving $\sum_{i=1}^n S_{\text{eff}}(\boldsymbol{X}_i, \boldsymbol{\theta}) = 0$, the semiparametric efficient estimator for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})^{\mathrm{T}}$ is obtained. The SE estimator for $\beta$ has the asymptotic estimation covariance matrix as follows.

$$\begin{aligned} & n\text{cov}(\widehat{\boldsymbol{\beta}}) \quad (2.5) \\ = & \Bigg( E\left[ \frac{m'_{\beta}(\boldsymbol{X}, \boldsymbol{\beta}_0) m'_{\boldsymbol{\beta}}(\boldsymbol{X}, \boldsymbol{\beta}_0)^T}{\sigma^2(\boldsymbol{X}, \boldsymbol{\gamma}_0)} \left\{ 1 + \frac{\mu_3^2}{\sigma^2(\boldsymbol{X}, \boldsymbol{\gamma}_0) E(C^2|\boldsymbol{X})} \right\} \right] \\ & - E\left\{ \frac{\mu_3 m'_{\boldsymbol{\beta}}(\boldsymbol{X}, \boldsymbol{\beta}_0) {\sigma^2}'^{T}_{\boldsymbol{\gamma}}}{\sigma^2(\boldsymbol{X}, \boldsymbol{\gamma}_0) E(C^2|\boldsymbol{X})} \right\} \left[ E\left\{ \frac{{\sigma^2}'_{\boldsymbol{\gamma}} {\sigma^2}'^{T}_{\boldsymbol{\gamma}}}{E(C^2|\boldsymbol{X})} \right\} \right]^{-1} E\left\{ \frac{\mu_3 m'_{\boldsymbol{\beta}}(\boldsymbol{X}, \boldsymbol{\beta}_0) {\sigma^2}'^{T}_{\boldsymbol{\gamma}}}{\sigma^2(\boldsymbol{X}, \boldsymbol{\gamma}_0) E(C^2|X)} \right\}^T \Bigg)^{-1}, \end{aligned}$$

where $\mu_3 = E(\epsilon^3|\boldsymbol{X})$. In the above equations, we use a subscript $_0$ to denote the true parameter value. Since the above covariance reaches Cramer-Rao lower bound (Tsiatis, 2006), the SE estimators are indeed efficient. They also verified that SE estimators for $\boldsymbol{\beta}$ are robust, even if the variance function $\sigma^2$ is misspecified.

#### 2.2.3. Weighting absolute centered external variable (WCEV)

In heteroscedastic error models, WLS methods use explanatory variables or a response variable to calculate weights. Unlike WLS methods, Çelik (2015) adopted WCEV for weights. WCEV is generally applicable to analyze data with various type of heteroscedasticity patterns including butterfly-distributed residuals and megaphone-shaped distributed residuals by applying optimal transformation to the external variable. Let $d$ be the external variable to calculate weights $w_i$ $(i = 1, 2, \cdots, n)$. In order to apply WCEV method, the data need to be sorted in ascending order by the response. Let $y_j$ be the response sorted in ascending order for $j = 1, \cdots, n$. First, we calculate the residuals $r_j$, $j = 1, 2, \cdots, n$ using OLS estimation. From the scatter plot of $r_j's$ against the newly sorted $\boldsymbol{X}$, we can figure out the heteroscedastic patten between the error and explanatory variable $\boldsymbol{X}$. Specifically, if the residuals are butterfly-distributed or right-megaphone-distributed then we set $d_j = 1, 2, \cdots, n$. If the residuals are left-megaphone-distributed, we set $d_j = n, \cdots, 2, 1$. Next, regarding the absolute value of residuals or squared residuals as a new response variable, build a regression model including d and $d^2$ as below.

$$|r| \sim \alpha_0 + \alpha_1 d + \alpha_2 d^2 \quad \text{or} \quad r^2 \sim \alpha_0 + \alpha_1 d + \alpha_2 d^2.$$

By differentiating the above regression with respect to $d$, we obtain the optimal nodal point $m$ for the possible $d$'s in the following.

$$m = -\frac{\alpha_1}{2\alpha_2}.$$

Now, we calculate the WCEV weights $w_j$'s for $j = 1, 2, \cdots, n$ in the following way.

$$w_j = \frac{1}{|d_j - m|^\delta}, \quad j = 1, 2, \cdots, n. \tag{2.6}$$

Using WCEV weights, weighted least squares estimation can be carried out. Here, optimum value of $\delta$ in (2.6) is estimated by maximizing the likelihood below.

$$L = \frac{n}{2}\left\{1 + \log(2\pi) + \log\left(\frac{SSR}{n}\right)\right\},$$

where SSR is the sum of squared residuals.

Çelik (2015) point out that WCEV is superior to other robust estimation methods because WCEV provides the less autocorrelated residulas, smaller CV and RMSE than other traditional estimation methods. Since WCEV does not violate regression assumptions and is applicable to data with various heteroscedasticity types, WCEV can be practically used in correcting heteroscedasticity problems. In particular, WCEV provides more robust estimation when the variance of residual is function of explanatory variable whose grid values are not smooth.

## 3. Data analysis

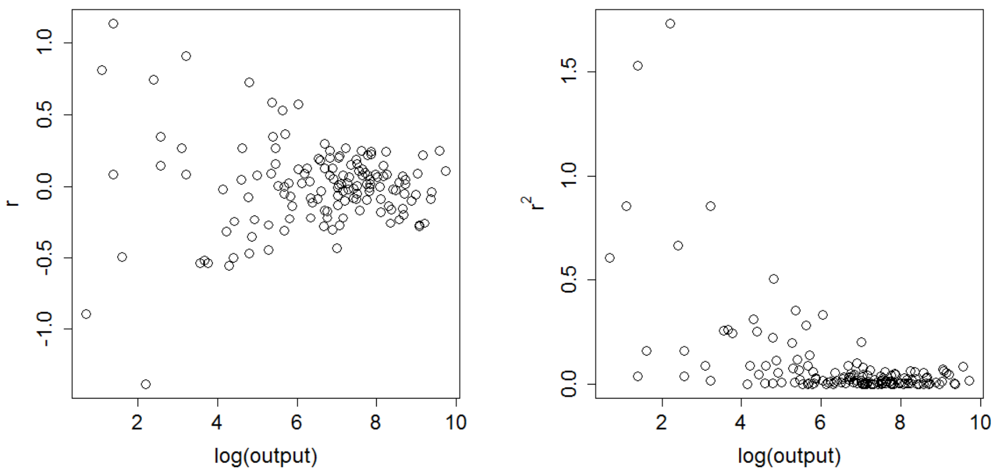
In this section, we analyze cost data of U.S. electricity producers using the regression estimation methods described in section 2. The data contains 159 observations and 8 variables about US electricity producers in 1955, which is currently available in R package AER. Among 159 observations, only first 145 observations are used for analysis except some observations from commonly owned firms. Christensen and Greene (1976) suggested the following generalized Cobb-Douglas cost function to model the scale of economy.

$$\log\left(\frac{C}{P_F}\right) = \beta_0 + \beta_1 \log Q + \beta_2 (\log Q)^2 + \beta_3 \log\left(\frac{P_K}{P_F}\right) + \beta_4 \log\left(\frac{P_L}{P_F}\right), \tag{3.1}$$

where $C$ is cost, $Q$ is output, $P_K$ is capital $P_L$ is labor and $P_F$ is fuel. According to economic theory, we can use the above equation as our regression mean function in (1.1) for our data. Let $Y = \log(C/P_F)$, $X_1 = \log Q$, $X_2 = \log(P_K/P_F)$ and $X_3 = P_L/P_F$. Then we consider the following model as a regression model.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_3 + \epsilon,$$

First, we perform OLS estimation for the above model we obtain residuals. In Figure 3.1, we provide the plot of residuals for OLS estimation against log(output), where we detect unequal variance according to covariates. For this reason, we can apply heteroscedastic error models to this data.




3.1: In the left, the plot of residuals of OLS estimation against log(output) is shown. In the right, the plot of residual squares of OLS estimation against log(output) is shown.

Note that the variance function $\mathrm{Var}(\epsilon|\boldsymbol{X})$ should be estimated in SE estimation and Bayesian estimation. We consider following variance function.

$$E(\epsilon^2|\boldsymbol{X}) = \exp(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\gamma}), \tag{3.2}$$

where $\boldsymbol{X} = (1, X_1, X_1^2, X_2, X_3)$, and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2, 0, 0)^{\mathrm{T}}$.

1. Box-Cox, Huber's $M$, bisquare, Yohai's MM

   WLS with the weight log(output) is performed by lm function in R. Huber's $M$, Bisquare, Yohai's MM are implemented using robustregH function and robustregBS function in robustreg package and lmrob function in robustbase package in R. For the Box-cox transformation, boxcox function in MASS package and bcPower function in car package are used. The estimated box-cox $\lambda$ is obtained to be 0.182.

2. Semiparametric efficient (SE) estimation

   Using residual squares of OLS estimation, the variance function is modeled. For the variance function in (3.2), we obtained the estimators for $\theta = \{\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \gamma_0, \gamma_1, \gamma_2\}$ which is represented in Table 3.1 and 3.2. To obtain $\widehat{C}$ and $\widehat{E[C^2|X]}$ in (2.4), we need to calculate $\widehat{\mu}_3$ and $\widehat{\mu}_4$. We estimated $\theta$ in the following procedures. First, we fit OLS model to obtain residuals ($r_{\text{OLS}}$'s). Next, regress $r_{\text{OLS}}^4$ on the dependent variable log(output) and its squared terms using OLS estimation. The fitted value of this model is used as $\widehat{\mu}_4$. We assigned sign$(r_{\text{OLS}}) \times \widehat{\mu}_4^{3/4}$ to $\widehat{\mu}_3$. R-package rootSolve is used to solve $\sum_{i=1}^n S_{\text{eff}}(\boldsymbol{X}_i, \boldsymbol{\theta}) = 0$. Also, we obtained the covariance matrix in (2.5) with R-package numDeriv.

3. Weighting absolute centered external variable (WCEV)

   Since the residuals obtained from OLS estimation have megaphone-distributed pattern in Figure 3.1, we apply WCEV with $d = \{1, 2, ..., n\}$. First, we sort data by ascending order of the response and carry out OLS regression to get the residuals $r_j$ $(j = 1, \cdots, n)$. Then regress $|r_j|$ on external variable $d$ and $d^2$. We have nodal point $m$=-0.639, $\hat{\delta}^{\text{MLE}}$=0.742 in (2.6) is obtained by using optim function in R-optim package.

4. Bayesian estimation

   As a prior distribution of the parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma})$, we assigned multivariate normal distribution where its mean equals to the OLS estimator and its variance matrix is a diagonal matrix having $100\times\text{SD}_{\text{OLS}}$ as the diagonal elements (Table 3.1 and 3.2). The variance matrix is set to be vague on purpose to accomplish reasonable acceptance rate for $\delta$, 23.4%. Note that the acceptance rate for $\boldsymbol{\beta}$ is 1, because $\boldsymbol{\beta}$ is sampled by gibbs sampling. In this paper, we produced 35,000 samples including 5,000 burn-in samples. Results in Table 3.1-3.2 and Figure 3.2-3.10 are reported after thinning (thin=3). We used R packages zoo, gridExtra, ggplot2, HDInterval, MASS, mvtnorm, coda to implement the algorithm and diagnosis convergence. We report trace plots and ACF plots are presented with ESS statistic by parameters. Figure 3.3-3.7 shows significantly low autocorrelation, while Figure 3.8-3.10 shows relatively high autocorrelation. Still, trace plots in Figure 3.3-3.10 imply sufficient state changes of samples.

**Table 3.1** The error variance model fit for variance function in (3.2)

| Methods | | Coefficients | | | Statistics |
|---|---|---|---|---|---|
| | | $\gamma_0$ | $\gamma_1$ | $\gamma_2$ | MSE |
| OLS | Estimate | 1.162 | -1.411 | 0.079 | 0.040 |
| | Std.error | 1.485 | 0.532 | 0.046 | |
| | t | 0.783 | -2.653 | 1.723 | |
| Bayesian | Estimate | 1.942 | -1.206 | 0.060 | 0.092 |
| | Std.error | 0.654 | 0.232 | 0.020 | |
| | t | 2.970 | -5.201 | 3.025 | |
| SE | Estimate | 0.667 | -0.791 | 0.028 | 0.033 |
| | Std.error | 0.947 | 0.299 | 0.022 | |
| | t | 0.704 | -2.644 | 1.279 | |

In Table 3.2, we report all estimations discussed in this paper. To compare those estimators, we report mean square errors (MSE), leave-one-out cross validation (LOOCV) and pseudo $R^2$ ($R^2$). LOOCV is calculated as follows.

$$\text{LOOCV} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2,$$

where $\widehat{y}_i$ is the predicted value from the estimation using data except $i^{th}$ observation ($i = 1, \cdots, n$). Because LOOCV tests all observations, LOOCV provides very stable results. Comparing OLS estimation, the standard errors of robust estimators are improved. Box-Cox transformation provides small standard error, but it is difficult to compare the coefficients of other methods directly, since the regression of Box-Cox transform is carried out using the transformed response. We compare Huber's $M$, bisquare and Yohai's MM estimations, which have similarities than others. Among those three estimators, bisquare estimators have smallest standard errors, but its LOOCV is bigger than other methods. Huber's $M$ shows the smallest LOOCV among three estimators. From the LOOCV values, we find that the Bayesian method is superior than other methods. However, Bayesian methods take longer time to obtain estimators than other methods. SE methods show the second best LOOCV and have smallest standard errors among all estimators. We can check SE estimator is indeed efficient from our data example. However, we have some trouble to find proper initial values to solve the efficient score functions numerically. When we apply WCEV methods, we obtained the smaller standard error than WLS estimation, while MSE and LOOCV of those estimations provide similar values.

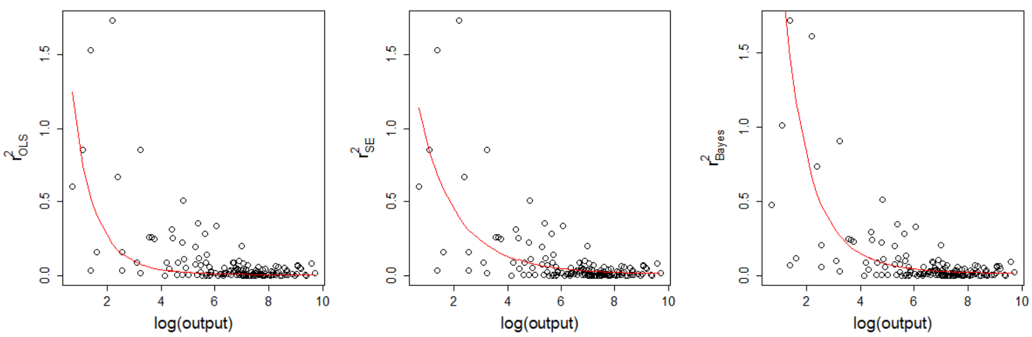



**Figure 3.2** Error variance estimation

**Table 3.2** Analysis results of heteroscedastic error models for the US Electricity data. Mean square errors (MSE), leave-one-out cross-validation (LOOCV) and Pseudo $R^2$ ($R^2$) are reported.

| Methods | | Coefficients | | | | | Statistics | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | MSE | LOOCV | $R^2$ |
| OLS | Estimate | -3.764 | 0.153 | 0.051 | 0.481 | 0.074 | 0.091 | 0.105 | 0.958 |
| | Std.error | 0.702 | 0.062 | 0.005 | 0.161 | 0.150 | | | |
| | t | -5.361 | 2.468 | 9.415 | 2.983 | 0.492 | | | |
| WLS | Estimate | -3.875 | 0.182 | 0.048 | 0.492 | 0.103 | 0.092 | 0.102 | 0.976 |
| | Std.error | 0.572 | 0.072 | 0.006 | 0.124 | 0.114 | | | |
| | t | -6.774 | 2.526 | 8.654 | 3.975 | 0.897 | | | |
| Box-Cox | Estimate | -1.949 | -0.241 | 0.068 | 0.363 | 0.064 | 0.103 | 0.123 | 0.969 |
| | Std.error | 0.452 | 0.040 | 0.003 | 0.104 | 0.097 | | | |
| | t | -4.311 | -6.050 | 19.695 | 3.496 | 0.663 | | | |
| Huber's $M$ | Estimate | -3.700 | 0.142 | 0.051 | 0.502 | 0.088 | 0.092 | 0.106 | 0.973 |
| | Std.error | 0.528 | 0.056 | 0.005 | 0.118 | 0.109 | | | |
| | t | -7.011 | 2.563 | 11.251 | 4.241 | 0.801 | | | |
| Bisquare | Estimate | -4.606 | 0.344 | 0.037 | 0.456 | 0.121 | 0.104 | 0.118 | 0.979 |
| | Std.error | 0.473 | 0.050 | 0.004 | 0.105 | 0.097 | | | |
| | t | -9.742 | 6.843 | 9.145 | 4.342 | 1.247 | | | |
| Yohai's MM | Estimate | -4.627 | 0.349 | 0.037 | 0.456 | 0.122 | 0.105 | 0.116 | 0.976 |
| | Std.error | 1.017 | 0.188 | 0.013 | 0.123 | 0.114 | | | |
| | t | -4.548 | 1.854 | 2.877 | 3.718 | 1.072 | | | |
| Bayesian | Estimate | -3.957 | 0.241 | 0.044 | 0.514 | 0.076 | 0.093 | 0.093 | 0.956 |
| | Std.error | 0.522 | 0.092 | 0.006 | 0.096 | 0.089 | | | |
| | t | -7.583 | 2.624 | 6.789 | 5.331 | 0.852 | | | |
| SE | Estimate | -3.707 | 0.206 | 0.046 | 0.542 | 0.042 | 0.092 | 0.100 | 0.958 |
| | Std.error | 0.387 | 0.050 | 0.004 | 0.076 | 0.073 | | | |
| | t | -9.578 | 4.114 | 12.942 | 7.095 | 0.584 | | | |
| WCEV | Estimate | -3.374 | 0.125 | 0.053 | 0.553 | 0.015 | 0.092 | 0.104 | 0.976 |
| | Std.error | 0.560 | 0.058 | 0.004 | 0.127 | 0.117 | | | |
| | t | -6.025 | 2.151 | 12.103 | 4.343 | 0.125 | | | |

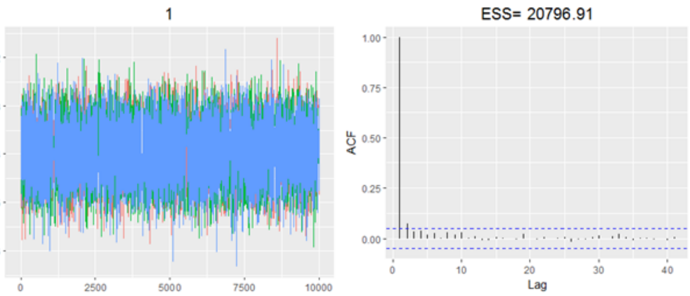



**Figure 3.3** Convergence diagnostics plot : $\beta_0$

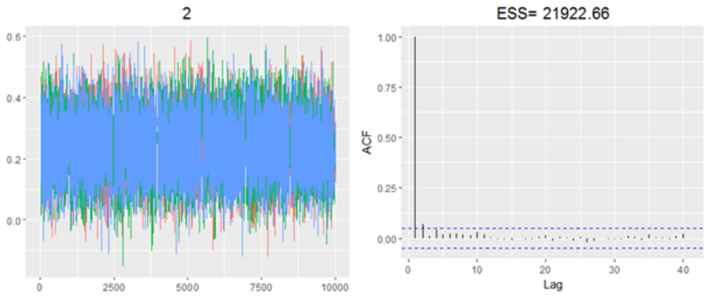



**Figure 3.4** Convergence diagnostics plot : $\beta_1$

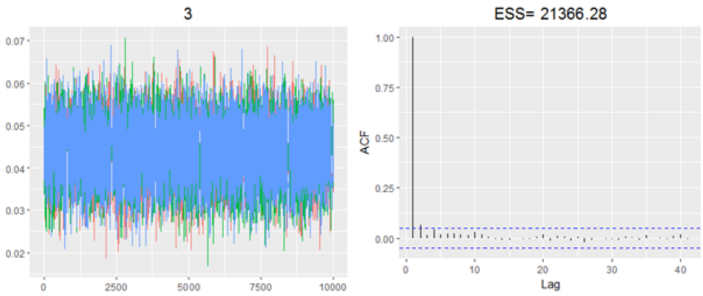



**Figure 3.5** Convergence diagnostics plot : $\beta_2$

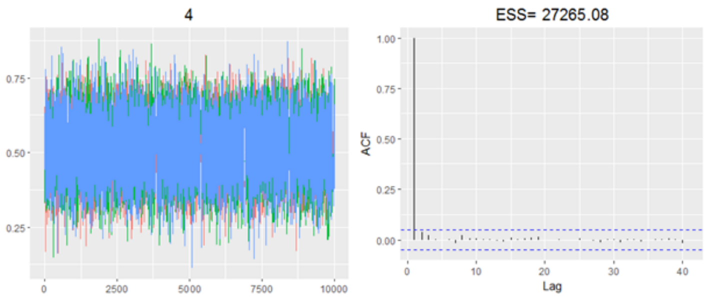



**Figure 3.6** Convergence diagnostics plot : $\beta_3$

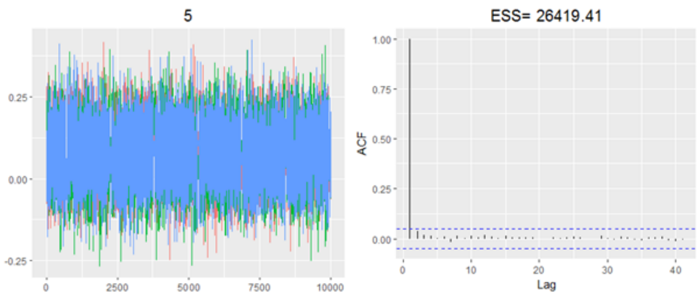



**Figure 3.7** Convergence diagnostics plot : $\beta_4$

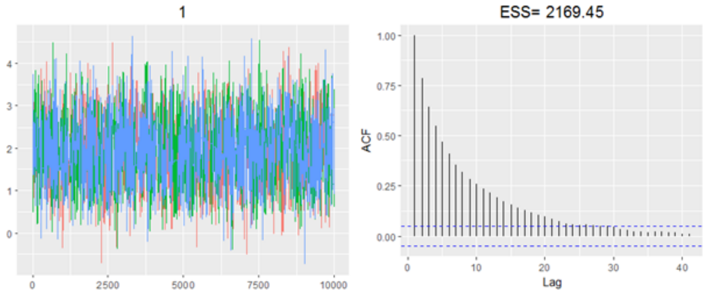



**Figure 3.8** Convergence diagnostics plot ; $\gamma_0$

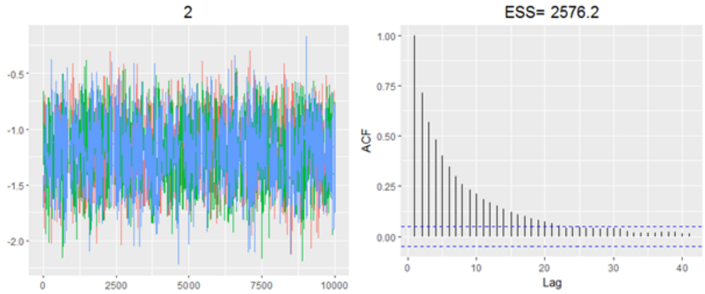



**Figure 3.9** Convergence diagnostics plot ; $\gamma_1$

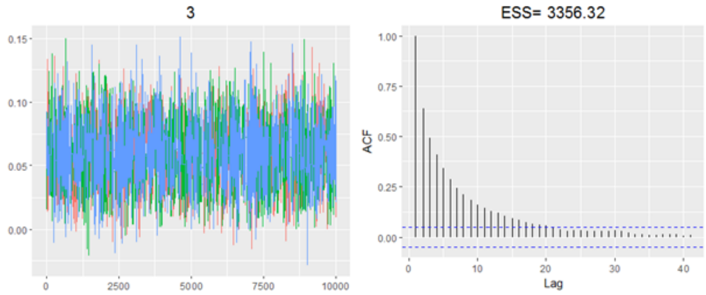



**Figure 3.10** Convergence diagnostics plot ; $\gamma_2$

## 4. Conclusion

In this paper, we revisit the well-known robust estimation methods. Box-Cox, Huber's $M$ and bisquare and Yohai's MM estimators show better results than WLS in terms of efficiency. However, LOOCV of Box-Cox, Bisquare and Yohai's MM estimators have greater LOOCV than WLS estimation from our data example. Only Huber's $M$ esimators has less LOOCV than WLS estimation. We also explained the recently introduced methods such as Bayesian, semiparametric estimation and WCEV estimation. We have found Bayesian estimation and semiparametric estimation perform better than other estimations. Also Bayesian estimation is useful because we can use highest posterior density (HPD) interval and Bayesian hypothesis test for inferences. However, Bayesian methods spend long time to obtain the results. The semiparametric estimation requires proper initial values to solve the efficient score function numerically. WCEV estimation carries out the weight method using absolute centered external variable and it improves efficiency comparing WLS. WCEV methods can be used easily because its concept is very simple and computationally inexpensive comparing Bayesian and semiparametric estimation. In addition to the methods presented in this paper, there are various ways of analyzing heteroscedastic data. We can also use regression estimation using both Bayesian methods and semiparametric methods in the similar way that Heo and Kim (2016) used. In particular, we can adopt a specific heteroscedastic error model on the structure of the response variable. For example, a beta regression model may be used, as Jang (2017) used when response variable takes a value between 0 and 1. It would be meaningful to apply the methods introduced in this paper to the data with the response variables having specific structure.

## References


Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B (Methodological)*, 211-252.

Çelik, R. (2015). Stabilizing heteroscedasticity for butterfly-distributed residuals by the weighting absolute centered external variable. *Journal of Applied Statistics*, **42**, 705-721.

Cepeda, E. and Gamerman, D. (2000). Bayesian modeling of variance heterogeneity in normal regression models. *Brazilian Journal of Probability and Statistics*, 207-221.

Cepeda, E. and Gamerman, D. (2005). Bayesian methodology for modeling parameters in the two parameter exponential family. *Revista Estadística*, **57**, 93-105.

Cepeda-Cuervo, E. (2015). Beta regression models: Joint mean and variance modeling. *Journal of Statistical Theory and Practice*, **9**, 134-145.

Cepeda-Cuervo, Ed and Núnez-Antón, V. (2007). Bayesian joint modelling of the mean and covariance structures for normal longitudinal data. *SORT*, **31**, 181-200.

Cepeda Cuervo, E. and Achcar, J. A. (2009). Regression models with heteroscedasticity using Bayesian approach. *Revista Colombiana de Estadstica*, **32**, 267-287.

Christensen, L. R. and Greene, W. H. (1976). Economies of scale in US electric power generation. *Journal of Political Economy*, **84**, 655-676.

Heo, S. and Kim, D. (2016). A comparative study in Bayesian semiparametric approach to small area estimation. *Journal of the Korean Data & Information Science Society*, **27**, 1433-1441.

Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 799-821.

Jang, E. (2017). An analysis of health-related quality of life using beta regression. *Journal of the Korean Data & Information Science Society*, **28**, 547-557.

Kim, M. and Ma, Y. (2012). The efficiency of the second-order nonlinear least squares estimator and its extension. *Annals of the Institute of Statistical Mathematics*, **64**, 751-764.

Rousseeuw, P. J. and Yohai, V. (1984). *Robust regression by means of S-estimators. In Robust and nonlinear time series analysis (pp. 256-272)*, Springer, New York.

Tsiatis, A. (2006). *Semiparametric theory and missing data*, Springer, New York.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, **15**, 642-656.