Robust Estimation Procedure for Autoregressive Models with Heterogeneity

A. Callens1 · Y.-G. Wang2 · L. Fu3 · B. Liquet1

Received: 7 January 2020 / Accepted: 24 August 2020 / Published online: 1 October 2020
© Springer Nature Switzerland AG 2020

Abstract
In environmental studies, regression analysis is frequently performed. The classical approach is the ordinary least squares method which consists in minimizing the sum of the squares of the residuals. However, this method relies on strong assumptions that are not always satisfied. In environmental data, the response variable often contains outliers and errors can be heteroscedastic. This can have significant effects on parameter estimation. To solve this problem, the weighted M-estimation was developed. It assumes a parametric function for the variance, and, estimates alternately and robustly, mean and variance parameters. However, this method is limited to the independent errors case, and is not applicable to time series data. Therefore, we introduce a new estimation procedure which adapts the weighted M-estimation to environmental time series data, while selecting optimal value for the tuning parameter present in the M-estimation. We compare the efficiency of our procedure on simulated data to other usual regression methods. Our estimation procedure outperforms the other methods providing estimates with lower biases and mean squared errors. Finally, we illustrate the proposed method using an air quality dataset from Beijing. This method has been implemented in the R package RlmDataDriven.

Keywords Heteroscedasticity · Model selection · Robust estimation · Temporal correlations

1 Introduction

In environmental modelling, pure homoscedasticity is uncommon. For example, residuals of hydrological [8], or air pollution models [27], are usually heteroscedastic. Ignoring this problem and performing an ordinary least squares method would result in regression parameters with biased covariance matrix and hence would lead to erroneous inference.

One method to deal with this biased covariance matrix is to use the White estimator [26] which provides a heteroscedasticity consistent covariance matrix. In case of a time series, the Newey-West estimator [18] can provide heteroscedasticity and an autocorrelation-consistent covariance matrix. Both of these methods account for heteroscedasticity but do not give information on the variability of the data generation process.

Another approach to cope with heteroscedasticity is to perform a weighted analysis, where the assumed underlying model is:

\[ y_i = x_i^T \beta + \sigma_i \epsilon_i, \]

with \( y_i \) the observed response, \( x_i \) the associated covariates, \( \beta \) the vector containing the parameters to be estimated, \( \epsilon_i \) the independently and identically distributed error terms with mean 0, and unknown symmetric distribution function and \( \sigma_i \) the term accounting for heteroscedasticity. Usually, a parametric function is assumed for this term [1, 4, 7, 9]. It can be a power function of the mean as proposed by Box et al. [3]: \( \sigma_i = \phi |x_i^T \beta|^\gamma \) or some functions of the covariates. Parameters in the variance function are not known and have to be estimated by the maximum likelihood method.
Estimation of parameters in models exhibiting heterogeneous variance is performed by an iterative procedure. A preliminary estimate of mean parameters is obtained by the least squares method. Residuals of this model are then used to estimate variance parameters. Finally, a weighted least squares method is performed with the estimated variance as weight. Unlike White and Newey-West methods, modeling heterogeneous variance enables one to get better estimates for the mean parameters and also to gain information on the variability of the data generation process [7, 9].

Such as least squares method, the estimation method for heteroscedastic models has a low breakdown point of 1/n, meaning that only one outlier in the observed response can have a large effect on the estimation of the mean parameters [21, 24, 28]. The estimation of the variance parameter is also affected as maximum likelihood methods are very sensitive to outliers [9, 22].

In practice, outliers are prevalent in the environmental dataset. They can be found in both response variable and covariates but, in this work, our interest lies only in outliers that are present in the response variable. In the presence of outliers, robust methods must be applied. They aim to produce reliable estimates that are not seriously affected by outliers, extreme values, or small deviations from model assumptions [13].

When both heteroscedasticity and outliers are present in regression analysis, one can use the method described by Carroll et al. [4]. This iterative method allows one to robustly estimate mean and variance parameters. For the mean parameters, they perform a weighted M-estimation with variance as weight. M-Estimation is a robust method which consists in minimizing a loss function that is slowly varying for abnormal residuals instead of squared residuals [25]. This loss function is controlled by a tuning parameter c which "regulates the amount of robustness" [11]. For the variance, they assume a parametric function for Huber’s loss function is 1.345 to ensure 95% asymptotic relative efficiency when data are normally distributed. More examples of loss functions can be found in Wang et al. [24, 25].

\[ \rho(u) = \begin{cases} \frac{1}{2}u^2 & \text{if } |u| \leq c \\ \frac{1}{2}c^2 & \text{if } |u| > c \end{cases} \]

Here, c is a tuning parameter chosen between 0 and 3, which controls the degree of robustness. Default value for Huber’s function is 1.345 to ensure 95% asymptotic relative efficiency when data are normally distributed.

Taking derivatives of (2) leads to the following estimating equation of \( \beta \):

\[ U(\beta) = \sum_{i=1}^{n} \left( \frac{x_i}{\hat{\sigma}_i} \right) \psi \left( \frac{y_i - x_i^T \beta}{\hat{\sigma}_i} \right) = 0, \]
where $\psi(x) = \min\{c, \max\{x, -c\}\}$ is the derivative of Huber's loss function.

To solve this estimating equation, one can rewrite $U(\beta)$ as:

$$U(\beta) = \sum_{i=1}^{n} x_i W_i r_i = 0,$$

dependent weights $W_i$, the robust estimator $\hat{\beta}$ can be obtained by the following formula:

$$\hat{\beta} = \left\{ \sum_{i=1}^{n} x_i W_i x_i^T \right\}^{-1} \left\{ \sum_{i=1}^{n} x_i W_i y_i \right\}.$$

An iterative approach is needed as $W_i$ is a function of $\beta$ and $\alpha$. This approach is derived from the pseudolikelihood approach and consists in fixing alternatively parameters of $\gamma$ and $\sigma$. There are two commonly used robust approaches to estimating the variance conditions, the robust estimator $\hat{\sigma}$ is obtained by the iterative procedure to be described below, this $\hat{\sigma}$ will be estimated from the variance function using the previous parameter estimates for $(\phi, \gamma, \beta)$. Here, $\eta_i$ represents the resulting error which should be close to $\sigma_i \xi_i$. We fit this augmented model with the optimal value of the tuning parameter to obtain the final estimate of $\beta$.

2.4 Accounting for Temporal Correlations

So far, we have only considered the independent model, we now need to incorporate the autoregressive process of order $p$ present in the error terms. We write $\epsilon_i$ as $\sum_{j=1}^{p} (\alpha_j \epsilon_{i-j}) + \xi_i$ where $\xi_i$ are independent errors and we rewrite the model (1) as:

$$y_i = x_i^T \beta + \sum_{j=1}^{p} \alpha_j \sigma_j \epsilon_{i-j} + \sigma_i \xi_i.$$

Because the $\epsilon_i$ are unobserved, we propose to use the Pearson residuals from the initial model (1), say, $\hat{\epsilon}_i$, and we now have the following linear model with roughly independent errors:

$$y_i = x_i^T \beta + \sum_{j=1}^{p} \alpha_j \hat{\sigma}_j \hat{\epsilon}_{i-j} + \eta_i,$$

where $(\hat{\sigma}_i \hat{\epsilon}_{i-1}, \hat{\sigma}_i \hat{\epsilon}_{i-2}, \ldots, \hat{\sigma}_i \hat{\epsilon}_{i-p})$ are the augmented additional covariates including $p$ lagged residuals, and $(\beta, \alpha_1, \ldots, \alpha_p)$ are the new parameters to be estimated including $p$ lag parameters, $\hat{\sigma}_i$ is an estimate of $\sigma_i$. In the iterative procedure to be described below, this $\hat{\sigma}_i$ will be estimated from the variance function using the previous parameter estimates for $(\phi, \gamma, \beta)$. Here, $\eta_i$ represents the resulting error which should be close to $\sigma_i \xi_i$. We fit this augmented model with the optimal value of the tuning parameter to obtain the final estimate of $\beta$.

2.5 The Estimation Procedure

The complete estimation procedure is summarized in the following algorithm (the covariates do not require to be scaled or normalized):
1. Obtain an initial robust estimate \( \hat{\beta}_0 \) assuming a constant variance \( g(x^T \beta, \gamma) = 1 \) and using M-estimation with the default value of \( c \) (\( rlm \) function).

2. By fixing \( \hat{\beta} = \hat{\beta}_0 \), the robust variance parameters \((\hat{\phi}, \hat{\gamma})\) are estimated with (4) and (5) respectively.

3. By fixing the variance parameters equal to their robust estimates, we update \( \hat{\beta} \) with (3).

4. To find the best tuning parameter, the steps 2–3 are repeated for a range of \( c \) values between 0 and 3. The best tuning constant is the one which minimizes the sum of the estimated variance of the regression parameters.

5. The model is fitted using the best value of the tuning constant \( \hat{c} \) from the steps 2–3. Then, temporal correlations are added by following the procedure described previously.

3 Numerical Studies

In this section, we investigate the performance of our procedure. We compare mean bias and mean squared errors of the estimates obtained by different methods such as least squares (\( \text{lm} \) function in R), generalized least squares method (\( \text{gls} \) function from the \( \text{nlme} \) package), M-estimation with \( c = 1.345 \) (\( \text{rlm} \) function from the \( \text{MASS} \) package), the weighted M-estimation (\( \text{whm} \) function from \( \text{rlmDataDriven} \) package) with \( c = 1.345 \), and our data-driven method (\( \text{rlmDD,het} \) function from \( \text{rlmDataDriven} \) package).

For one simulation, we generate a multivariate normal dataset \((n = 500)\) using the model (1). In our case, \( x_i^T \beta = \beta_0 + \beta_1 x_{i1} \) where we fix the value of \( \beta_0 \) and \( \beta_1 \) to 10 and \( x_{1i} \) comes from a uniform distribution on \((0, 1)\). For \( \sigma_i \), we test two functions: the power function \( \sigma_1 = |x_i^T \beta|^\gamma \) with \( \gamma = 0.3 \) and the exponential function \( \sigma_2 = e^{\gamma|x_i^T \beta|} \) with \( \gamma = 0.05 \). For the term \( \epsilon_i \), we consider two cases: (i) an autoregressive process of order 1 with \( \alpha = 0.5 \) and (ii) an autoregressive process of order 2 with \( \alpha_1 = 0.5, \alpha_2 = 0.3 \). These processes can be written as follows: \( \epsilon_i = \sum_{j=1}^p \alpha_j \epsilon_{i-j} + \xi_i \) where \( \xi_i \) are independent and normally distributed errors following a standard normal distribution \( N(0, 1) \).

Two methods of contamination are considered. On the first case, a proportion of \( \lambda \) data are randomly selected as outliers. A fixed value of 10 is then added to contaminate each of these \( y_i \) values. This way of contamination is standard in literature [14].

On the second case, we sample randomly a proportion \( \lambda \) of the 50% highest values of \( x_{1i} \). For the selected \( \lambda \%), we add a value of 10 to the corresponding \( y_i \) values. By using this contamination method, we imitate a process where the chances of getting an outlier increase with the value of the explanatory variable \( x_{1i} \). In both cases, several contamination rates are tested: \( \lambda = 0\%, 5\%, 10\% \).

Table 1 shows the means and the standard errors of the estimate bias computed across all the simulations for the exponential function. The results obtained with the power function are included in the Online Supplementary Material (Table S-1).

For both cases of contamination and variance functions, average value of the data-dependent tuning constant decreases as contamination rate increases. This tuning constant has the expected behavior: as the proportion of outliers becomes larger, more values should be considered as outliers therefore the tuning constant is smaller.

For random contamination (case 1), only the estimates of \( \beta_0 \) are affected by the outliers. The mean bias observed for \( \beta_0 \) increases with the proportion of outliers for every method. However, it is less important for robust methods, especially for the proposed method. Indeed, it yields estimators with lower mean bias when outliers are present. In addition, the proposed method performs better when the autoregressive process is of order 2.

For the second case of contamination, the estimates of both \( \beta_0 \) and \( \beta_1 \) are affected by the outliers. Such as the first case, the mean biases increase with the contamination rate for every method but this phenomenon is less important for the robust methods. Even if the biases are still present, they are lower when the proposed method is used.

In terms of mean squared errors, this method provides estimates with lowest MSE for the first case of contamination. Differences in efficiency with the other methods augment with the contamination rate, reaching MSE values for \( \beta_0 \) nearly fivefold lower than the ones obtained by the least squares method (Fig. 1a).

For the second case of contamination, the weighted M-estimation seems to perform slightly better than the proposed method for the estimate of \( \beta_0 \). However, the MSE values obtained by our method are still really inferior to the ones obtained by the least squares method. For \( \beta_1 \), our method yields estimates with lowest MSE when contamination is greater than 5%.

The same observations can be drawn from the figure of the power variance function available in the Supplementary Material (Fig. S-1). In this framework, our method clearly outperforms the other approaches in terms of MSE especially for \( \beta_1 \) in case 1.

In terms of computation time, \( \text{rlmDD,het} \) (the R function for the proposed method) took an average 0.08 s to estimate the coefficients for one simulation. The \( \text{gls} \) and \( \text{whm} \) functions took an average 0.01 s. Finally, the \( \text{rlm} \) and \( \text{lm} \) functions took an average less than 0.001 s. The computation time is slightly longer with the \( \text{rlmDD,het} \) function because the function performs several estimations with different values of the tuning constant \( c \) in order to find its optimal value.
Table 1  Mean biases and associated standard errors of the estimated parameters obtained by several regression methods for the exponential variance function. Results based on 500 replications

| Case | $\sigma_i = e^{\gamma |x_i^T \beta|}$, $\gamma = 0.05$, AR(1), $\alpha = 0.5$ | $\lambda = 0\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ |
|------|-----------------------------------------------------------|----------------|----------------|----------------|
|      | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_0$ | $\hat{\beta}_1$ |
| Least square | 0.01(0.01) | -0.01(0.02) | 0.49(0.01) | 0.02(0.02) | 1.01(0.02) | -0.04(0.03) |
| M-Estimatin | 0.01(0.01) | -0.01(0.02) | 0.2(0.01) | 0.03(0.02) | 0.46(0.01) | 0.04(0.02) |
| Generalized least square | 0.01(0.01) | -0.01(0.02) | 0.49(0.01) | -0.01(0.02) | 1.01(0.02) | -0.04(0.03) |
| Weighted M-estimation | -0.01(0.01) | -0.01(0.02) | 0.09(0.01) | 0.06(0.02) | 0.21(0.01) | 0.16(0.02) |
| Proposed method | 0.01(0.01) | -0.01(0.02) | 0.06(0.01) | 0.02(0.02) | 0.24(0.01) | 0.02(0.02) |
| $\hat{c}$ | 2.08 | 1.04 | 0.75 |

| Case | $\sigma_i = e^{\gamma |x_i^T \beta|}$, $\gamma = 0.05$, AR(2), $\alpha_1 = 0.5$, $\alpha_2 = 0.3$ | $\lambda = 0\%$ | $\lambda = 5\%$ | $\lambda = 10\%$ |
|------|-----------------------------------------------------------|----------------|----------------|----------------|
|      | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_0$ | $\hat{\beta}_1$ |
| Least square | 0.04(0.02) | -0.01(0.02) | 0.5(0.02) | 0.02(0.03) | 1.01(0.03) | 0.03(0.03) |
| M-Estimatin | 0.04(0.02) | -0.02(0.02) | 0.27(0.02) | 0.04(0.02) | 0.6(0.03) | 0.07(0.03) |
| Generalized least square | 0.04(0.02) | -0.01(0.02) | 0.5(0.02) | 0.03(0.03) | 1(0.03) | 0.03(0.03) |
| Weighted M-estimation | 0.05(0.02) | -0.02(0.03) | 0.16(0.02) | 0.09(0.02) | 0.34(0.03) | 0.27(0.03) |
| Proposed method | 0.03(0.02) | 0.01(0.02) | 0.09(0.02) | 0.01(0.02) | 0.3(0.03) | 0(0.03) |
| $\hat{c}$ | 1.69 | 1.04 | 0.77 |

4 Application to Air Quality Data

4.1 Context of the Application

We apply the proposed robust procedure to analyze fine particle matter (PM$_{2.5}$) concentrations from the US embassy situated in Beijing, China. Such as other China megacities, Beijing has been suffering from chronic air pollution [5]. The main constituents of this pollution are suspended particle matters under 2.5 µm of diameter, widely known as PM$_{2.5}$ [16]. Concentration of PM$_{2.5}$ is highly variable and depends on sources of emission, secondary chemical generation processes and meteorological conditions. According to numerous studies [16, 19, 29], these suspended particles affect climate, visibility, and human health in many ways.
Given the severity of the pollution and the potential hazardous effects, China’s State Council aimed to reduce the PM$_{2.5}$ pollution by at least 25% for the period 2012–2017. Our objectives are (i) to show the efficiency of our procedure on a real dataset compared with a conventional estimation method (least squares method), and (ii) to analyze the potential effect of the decision of China’s State Council on the concentration of PM$_{2.5}$. Our method is
highly desirable in this case study since the variance of PM$_{2.5}$ concentrations is not likely to be homogeneous \cite{27} and outliers may be present.

### 4.2 Regression Analysis

The hourly data come from a previous study lead by Liang et al. \cite{16} (Fig. 2). PM$_{2.5}$ concentrations were taken at the US Embassy of Beijing and meteorological measurements at the Beijing Capital Airport. Both time series covers the period from 1 January 2010 to 31 December 2014.

To evaluate changes in PM$_{2.5}$ concentration after the decision of China’s State Council, we created two new variables: policy and time policy. Both variables take the value 0 before 2012; however, policy takes the value 1 after 2012 to detect any shifts in the intercept and policy time takes the value of the time lapsed in days after 2012 to test and quantify the trend after that decision.

We modelled PM$_{2.5}$ concentration as a linear combination of the following covariates: dew point (°C), temperature (°C), atmospheric pressure (hPa), combined wind direction (3 factors), cumulated wind speed (m/s), cumulated hour of snow (mm), cumulated hour of rain (mm), time policy, policy, and seasonal patterns with $\sin$ and $\cos$ functions for the 3- and 2-year and 6-, 4-, and 3-month cycles. To facilitate the interpretation of the estimates across the covariates, the following numerical covariates have been standardized: dew point, temperature, atmospheric pressure, cumulated wind speed, cumulated hour of snow, cumulated hour of rain.

The response variable PM$_{2.5}$ is a concentration; therefore, a transformation was needed in order to avoid predictions lower than zero with the regression analysis. Two transformations were tested on the dataset: square-root and logarithm transformation (Fig. 2). For both transformations, re-transformating fitted values in original scale. Although logarithm transformation is the classical transformation for air quality data, we chose the square root transformation as it yielded slightly lower errors (Supporting information, Table S-3).

The model is fitted with least squares method by the `lm` function from R statistical software \cite{20}. The residuals versus fitted value plot (Fig. 3) indicate the presence of heteroscedasticity with larger residuals as fitted values increase. This heteroscedasticity does not lead to biased estimators but to estimators with biased covariance matrix. This could result in underestimation of standard errors, erroneous Z-values, and therefore erroneous hypothesis tests. The normal probability plot in Fig. 3 reveals that distribution of residuals is skewed, indicating the presence of outliers which may have influenced the estimation of regression parameters. Moreover, temporal correlations in the residuals have been found using the ACF and pACF plots.

In this present case, our method is highly desirable as we have the presence of heteroscedasticity, temporal correlations, and outliers.

### 4.3 Robust Regression Analysis with the Proposed Method

We fitted the same regression model to the data with the proposed method. This method has been implemented in a R function: `rlmDD_het`, available in the package `RlmDataDriven`. The R code for the presented analysis can be found in the Supporting Information file.

In the literature, we did not find previous papers or indications on how to model variance of PM$_{2.5}$ concentration; consequently, we used common variance functions for the analysis such as power or exponential functions. Hereafter, we present the result obtained by considering $\sigma_i = \phi e^{r_{i}^{1.5} \hat{\beta}}$.

First, our method chooses the best tuning constant by testing a range of values for the tuning parameter between 0 and 3. As stated earlier, the best value of $c$ will be the value that minimizes the sum of the variance of the regression parameters. For the PM$_{2.5}$ data, optimal value of the tuning parameter is found around 1.5.

Then, we use ACF and pACF plots of the robustified residuals to determine the order of the autoregressive process. The robustified residuals are defined as $\psi((y_i - \hat{x}_i^T \hat{\beta}) / \hat{\sigma}_i)$. Two significant lags are found in the pACF plot; therefore, we consider that $p = 2$ and we add two lagged terms to the regression model. The lagged terms are:

$$\alpha_1 \hat{\sigma}_i \hat{\epsilon}_{i-1} + \alpha_2 \hat{\sigma}_i \hat{\epsilon}_{i-2} + \eta,$$

where $\hat{\sigma}_i \hat{\epsilon}_{i-1}, \hat{\sigma}_i \hat{\epsilon}_{i-2}$ are lagged terms built from the initial model estimated with the best tuning constant ($\hat{c} = 1.5$) and $\eta$ is the assumed independent error term. The term $\hat{\epsilon}$ corresponds to Pearson residuals of the initial model and $\hat{\sigma}$ is the estimated variance function.

The normal probability plot (Fig. 3) of the residuals clearly illustrate that our robust procedure has taken care of outliers successfully. The residuals versus fitted values plot (Fig. 3) seems to indicate that heteroscedasticity has been lowered.

The results of both methods are listed in Table 2. In this table, the covariance matrix of the regression parameters obtained by least squares method has been estimated with the NeweyWest function (sandwich package) which gives a heteroscedasticity and autocorrelation consistent estimation of the covariance. This estimation was necessary to obtain corrected standard errors as the residuals exhibited heteroscedasticity and temporal correlations.

Our estimation method drastically reduces the variance of the parameters. It is worth noting that the covariates policy and time policy are not significantly different from 0.
**Fig. 2** Time series of the PM$_{2.5}$ concentration with different transformations. The first panel shows the raw data, the second and last ones show the square-root and logarithm transformation respectively.

**Fig. 3** Regression diagnostic plots for the least squares model (upper panel) and the proposed method (lower panel). In the residuals vs fitted values plot for the method, 6 residuals lie outside the plot area.
Yearly cycles have a significant effect on the PM concentration due to cyclical variations in meteorological conditions. In the least squares method contrary to our robust method. Years after 2011 are characterized by a positive shift in the intercept. However, coefficient of the variable time policy is slightly negative. After 1 year, this coefficient outweighs the positive shift in the intercept, meaning that the PM$_{2.5}$ concentration decreases slightly compared with that of years before 2012.

Finally, we can see in Table 2 that lagged terms are significant. This indicates that the two previous terms contribute significantly to the output and were, therefore, necessary to consider.

5 Discussion

This method is data-dependent by the optimal choice of the tuning constant and it incorporates temporal correlations by adding lagged terms in the covariates. The numerical study showed that this procedure outperforms the other usual regression methods when data are contaminated by providing more precise estimates for the mean parameters. In the application with the PM$_{2.5}$ concentration dataset, we proved that our method results in estimates with significantly lower variance compared with the ones obtained by least squares estimation, leading to better hypothesis testing. This method is well suited for environmental dataset due to the frequent presence of heterogeneity and outliers. Hyslop et al. [12] utilized Thiel-Sen robust regression to evaluate long-term trends in aerosol concentrations via the historical PM$_{2.5}$ element measurements. Van Donkelaar et al. [23] used a geographically weighted regression (GWR) statistical model to represent bias of fine PM$_{2.5}$ concentrations over North America. Knibbs et al. [15] utilized the land-use regression (LUR) to estimate PM$_{2.5}$ at continental scale and explained the most spatial variability in PM$_{2.5}$ in Australia. These three methods did not consider the heterogeneity and autoregressive errors. In the future, the proposed method could be generalized to these three models. Furthermore, the proposed method could be extended to model time series of counts by using a link function such as the

Table 2 Parameter estimates ($\beta$), their standard errors, and z-values for the least squares and proposed method. The critical region of the significance test is of the full model $|z| > 1.96$

| Parameter      | Least squares method | Proposed method with $\hat{c} = 1.5$ and $\hat{\sigma}_i = 1.03^{0.131|\beta|}$ |
|----------------|----------------------|---------------------------------------------------------------------------------------|
|                | Estimate             | Std. error | Z-value | Estimate | Std. error | Z-value |
| Intercept      | 9.378                | 0.157      | 59.765  | 8.784    | 0.018      | 493.763  |
| Dew point      | 3.214                | 0.122      | 26.446  | 2.953    | 0.011      | 268.731  |
| Temperature    | -3.978               | 0.148      | -26.945 | -3.416   | 0.013      | -260.880 |
| Pressure       | -0.882               | 0.149      | -5.909  | -0.877   | 0.011      | -77.211  |
| Chwd NE        | -1.455               | 0.104      | -13.952 | -1.117   | 0.021      | -52.428  |
| Chwd NW        | -1.716               | 0.099      | -17.284 | -1.400   | 0.018      | -77.689  |
| Chwd SE        | 0.512                | 0.081      | 6.306   | 0.666    | 0.019      | 35.685   |
| Iws            | -0.577               | 0.058      | -9.888  | -0.410   | 0.005      | -89.071  |
| Is             | -0.086               | 0.054      | -1.575  | -0.027   | 0.008      | -3.475   |
| Ir             | -0.453               | 0.039      | -11.677 | -0.380   | 0.004      | -88.028  |
| Policy         | 0.362                | 0.343      | 1.058   | 0.222    | 0.023      | 9.464    |
| Time policy    | $-7 \times 10^{-4}$  | $5 \times 10^{-4}$ | -1.373  | $-6 \times 10^{-4}$ | $3 \times 10^{-5}$ | -18.052  |
| Sin 3 years cycle | 0.547               | 0.174      | 3.155   | 0.509    | 0.010      | 48.993   |
| Cos 3 years cycle | -0.340              | 0.157      | -2.164  | -0.238   | 0.010      | -22.878  |
| Sin 2 years cycle | 0.206               | 0.133      | 1.549   | 0.345    | 0.008      | 43.472   |
| Cos 2 years cycle | 0.084               | 0.123      | 0.683   | 0.104    | 0.009      | 11.361   |
| Sin 6 months cycle | -0.342              | 0.123      | -2.776  | -0.355   | 0.008      | -42.542  |
| Cos 6 months cycle | -0.365              | 0.119      | -3.069  | -0.425   | 0.008      | -51.631  |
| Sin 4 months cycle | -0.184              | 0.120      | -1.539  | -0.038   | 0.008      | -4.722   |
| Cos 4 months cycle | -0.035              | 0.121      | -0.287  | -0.084   | 0.008      | -10.509  |
| Sin 3 months cycle | 0.032               | 0.120      | 0.263   | 0.047    | 0.008      | 5.947    |
| lag1           | /                    | /          | /       | 0.746    | 0.005      | 162.733  |
| lag2           | /                    | /          | /       | 0.102    | 0.004      | 23.260   |
generalized linear model [17]. Concerning the variance, it would be helpful to consider more flexible approaches for the estimation of the parameters and to provide guidance for choosing the most appropriate variance function for a dataset.

Supplementary Material

Online supplementary material includes:

– Table S-1: Mean biases and associated standard errors of the estimated parameters obtained by several regression methods for the power variance function.
– Table S-2: Mean Absolute Error and Root mean squared error for different transformations of PM$_{2.5}$ and different models.
– Figure S-1: Mean squared errors of the estimated parameters obtained by several regression methods for the power variance function.
– Page 4: Information on the dataset with downloading link and indications on the different steps to tidy the dataset before analyzing it.
– The R script used to obtain the results in the application section.

Acknowledgements The work was carried out during the visit of Aurelien Callens to the School of Mathematical Science, Queensland University of Technology, Brisbane, Australia.

Funding This research was partially funded by the Australian Research Council project (DP160104292). Liya Fu’s research was supported by the National Natural Science Foundation of China (11871390), the Fundamental Research Funds for the Central Universities (No. xjj2017180), the Natural Science Basic Research Plan in Shaanxi Province of China (2018JQ1006).

References

1. Bianco, A., Boente, G., Di Rienzo, J. (2000). Some results for robust GM-based estimators in heteroscedastic regression models. *Journal of Statistical Planning and Inference*, 89(1-2), 215–242.
2. Bickel, P.J. (1978). Using residuals robustly I: Tests for heteroscedasticity, nonlinearity. *The Annals of Statistics*, 6(2), 266–291.
3. Box, G.E., & Hill, W.J. (1974). Correcting inhomogeneity of variance with power transformation weighting. *Technometrics*, 16(3), 385–389.
4. Carroll, R.J., & Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. *The Annals of Statistics*, 10(2), 429–441.
5. Chan, C.K., & Yao, X. (2008). Air pollution in mega cities in China. *Atmospheric Environment*, 42(1), 1–42.
6. Croux, C. (1994). Efficient high-breakdown M-estimators of scale. *Statistics & Probability Letters*, 19(5), 371–379.
7. Davидian, M., & Carroll, R.J. (1987). Variance function estimation. *Journal of the American Statistical Association*, 82(400), 1079–1091.
8. Evin, G., Kavetski, D., Thyer, M., Kuczera, G. (2013). Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration. *Water Resources Research*, 49(7), 4518–4524.
9. Giltinan, D.M. (1983). Bounded influence estimation in heteroscedastic linear models. Ph.D. thesis, Citeseer.
10. Huber, P.J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 1(5), 799–821.
11. Huber, P.J. (1981). Robust statistics. New york, Chichester, Brisbane. Toronto, Singapore: John Wiley & Sons Ltd.
12. Hyslop, N.P., Trzepla, K., White, W.H. (2015). Assessing the suitability of historical pm2.5 element measurements for trend analysis. *Environmental Science & Technology*, 49(15), 9247–9255.
13. Jiang, Y., Wang, Y.G., Fu, L., Wang, X. (2019). Robust estimation using modified huber’s functions with new tails. *Technometrics*, 61(1), 111–122.
14. Jiang, Y., Wang, Y.G., Fu, L., Wang, X. (2019). Robust estimation using modified huber’s functions with new tails. *Technometrics*, 61(1), 111–122.
15. Knibbs, L.D., van Donkelaar, A., Martin, R.V., Bechle, M.J., Brauer, M., Cohen, D.D., Cowie, C.T., Dirgawati, M., Guo, Y., Hanigan, I.C., et al. (2018). Satellite-based land-use regression for continental-scale long-term ambient pm2.5 exposure assessment in Australia. *Environmental Science & Technology*, 52(21), 12445–12455.
16. Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., Huang, H., Chen, S.X. (2015). Assessing Beijing’s PM2.5 pollution: Severity, weather impact, APEC and winter heating. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2182), 20150257.
17. McCullagh, P., & Nelder, J.A. (1989). Generalized Linear Models. vol. 37, CRC press.
18. Newey, W.K., & West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708. http://www.jstor.org/stable/1913610.
19. Pope, C.A. III., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurston, G.D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama*, 287(9), 1132–1141.
20. R Core Team (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
21. Rousseauw, P.J., & Leroy, A.M. (2005). Robust Regression and Outlier Detection. vol. 509, John wiley & sons.
22. Stefanski, L.A., Carroll, R.J., Ruppert, D. (1986). Optimally bounded score functions for generalized linear models with applications to logistic regression. *Biometrika*, 73(2), 413–424.
23. Van Donkelaar, A., Martin, R.V., Spurr, R.J., Burnett, R.T. (2015). High-resolution satellite-derived pm2.5 from optimal estimation and geographically weighted regression over north america. *Environmental Science & Technology*, 49(17), 10482–10491.
24. Wang, N., Wang, Y.G., Hu, S., Hu, Z.H., Xu, J., Tang, H., Jin, G. (2018). Robust regression with data-dependent regularization parameters and autoregressive temporal correlations. *Environmental Modeling & Assessment*, 23(6), 1–8.
25. Wang, Y.G., Lin, X., Zhu, M., Bai, Z. (2007). Robust estimation using the Huber function with a data-dependent tuning constant. *Journal of Computational and Graphical Statistics*, 16(2), 468–481.
26. White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838. http://www.jstor.org/stable/1912934.
27. Wu, E.M.Y., & Kuo, S.L. (2012). Air quality time series based garch model analyses of air quality information for a total quantity control district. *Aerosol and Air Quality Research, 12*(3), 331–343.

28. Zhao, J., & Wang, J. (2009). Robust testing procedures in heteroscedastic linear models. *Communications in Statistics–Simulation and Computation*, 38(2), 244–256.

29. Zheng, M., Salmon, L.G., Schauer, J.J., Zeng, L., Kiang, C.S., Zhang, Y., Cass, G.R. (2005). Seasonal trends in PM2.5 source contributions in Beijing, China. *Atmospheric Environment*, 39(22), 3967–3976.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.