IJAMML 6:1 (2017) 27-34 March 2017
ISSN: 2394-2258
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/ijamml_7100121817

ROBUST ESTIMATION FOR LINEAR ERRORS-IN-VARIABLES MODELS WITH HOMOSCEDASTIC MEASUREMENT ERRORS

Cuiping Guo[a], Junhuan Peng[b] and Chuantao Li[a]

[a]School of Science, China University of Geosciences (Beijing), Beijing, 100083 P. R. China
[b]School of Land Science and Technology, China University of Geosciences (Beijing), Beijing, 100083, P. R. China

Abstract

Linear errors-in-variables (EIV) models with heteroscedastic measurement errors are widely adopted in applied sciences. The linear EIV model estimators, however, can be highly biased by gross errors. This paper focuses on robust estimation in linear homoscedastic EIV models. Robust estimators of the linear homoscedastic EIV models are derived from M-estimators and Lagrange multiplier method.

Keywords: errors-in-variables, homoscedastic measurement errors, robust estimation, total least squared estimators, M-estimators.

*Corresponding author.
E-mail address: gcp@cugb.edu.cn (Cuiping Guo).

2010 Mathematics Subject Classification: 62F10.
Submitted by Jianqiang Gao.
Received February 10, 2017

1. Introduction

The linear regression model is a vital tool for statistical analysis in many areas such as signal processing, metrology, and engineering. In most cases, the independent variable is assumed to be known exactly, while the response variable is considered to have measurement errors. Usually, a least-squares procedure is employed to get the most favorable linear unbiased estimation of the parameters. However, the measured independent variable may contain errors due to the influence of instrument error, observation condition and human factors. The measured errors existing in independent variable disturb the optimal properties of least squares estimators, and result in the bias in the estimators for the slope and intercept. Suppose that there are true unobserved variables (ξ_i, η_i) that satisfy the linear relationship

$$\eta_i = \beta_1 + \beta_2 \xi_i \quad (i = 1, \cdots, n). \tag{1}$$

We observe (x_i, y_i) which are the true variables plus the additive measurement errors $(\delta_i, \varepsilon_i)$, namely,

$$x_i = \xi_i + e_{x_i}, \; y_i = \eta_i + e_{y_i} \quad (i = 1, \cdots, n). \tag{2}$$

Model (1)-(2) is called a linear errors-in-variables (EIV) model. In this paper, we assume that $\delta_1, \cdots, \delta_n; \varepsilon_1, \cdots, \varepsilon_n$ are independent error variables and $\delta_i \sim (0, \sigma^2_{e_{x_i}})$, $\varepsilon_i \sim (0, \sigma^2_{e_{y_i}})$.

The author first associated with the linear EIV models is Adcock [1]. However, Adcock only considered equal error variances $(\sigma^2_{e_{x_1}} = \cdots = \sigma^2_{e_{x_n}} = \sigma^2_{e_{y1}} = \cdots = \sigma^2_{e_{yn}})$. Adcock's work was later extended by Kummel [9] and instead of assuming equal error variances, he assumed that $\sigma^2_{e_{x_1}} = \cdots = \sigma^2_{e_{x_n}} = \sigma^2_x$, $\sigma^2_{e_{y1}} = \cdots = \sigma^2_y$, and the ratio $\lambda = \sigma^2_x / \sigma^2_y$ was known instead. Later, the linear EIV model has been widely discussed in quite a number of publications (Cheng and Van Ness [2];

Fuller [4]). There are many approaches to deal with the linear EIV model. The method of moments approach is considered in detail in Gillard and Iles [5]. The vast majority of the papers available on the linear EIV model have adopted a maximum likelihood approach to estimate the parameters (Wong [13]; Gillard and Iles [5]). Cheng and Rui [3] discussed maximum likelihood estimators in heteroscedastic simple linear EIV models with normal errors. Total least squares is a method of estimating the parameters of a linear EIV model and was introduced by Golub and Van Loan [6], which is frequently cited in the computational mathematics and engineering literature. Furthermore, in statistical and geodetic literature, many authors focused on the solution of EIV models using the weighted total least-squares (WTLS) methods (Schaffrin et al. [11]; Schaffrin and Wieser [12]; Mahboub et al. [10]). However, similar to the traditional EIV model, the estimated parameters of the linear EIV model with heteroscedastic measurement errors are affected or even severely distorted when observed variables are contaminated by outliers, and so robust estimation is required for a linear EIV model. The most common method of robust regression is M-estimation by Huber [7].

In this paper, we investigate robust estimation of the linear heteroscedastic EIV model by M-estimation and Lagrange multiplier method. The structure of this paper is as follows. Section 2 introduces robust estimators for the linear heteroscedastic EIV model when the independent and response variables are contaminated by gross errors. The paper is concluded in Section 3.

2. Robust Estimators for Linear Heteroscedastic EIV Models

Suppose the true values of response variable and independent variables are linearly related as

$$x := [x_1, \cdots, x_n]^T \text{ and } y := [y_1, \cdots, y_n]^T.$$

Then these form the relationship

$$y - e_y = \beta_1 \cdot 1_n + \beta_2 \cdot (x - e_x), \tag{3}$$

where $1_n = [1, \cdots, 1]^T$ and

$$\begin{bmatrix} e_y \\ e_x \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma_o^2 \begin{bmatrix} Q_y & 0 \\ 0 & Q_x \end{bmatrix} \right), \; P_y = Q_y^{-1}, \; P_x = Q_x^{-1}.$$

In this paper, we just studied the case, in which the weight matrices P_y and P_x are diagonal matrices defined as

$$P_y = \begin{pmatrix} r_1 & & & \\ & r_2 & & \\ & & \ddots & \\ & & & r_m \end{pmatrix}, \; P_x = \begin{pmatrix} S_1 & & & \\ & S_2 & & \\ & & \ddots & \\ & & & S_n \end{pmatrix}.$$

The weighted total least squares principle is to minimize the objective function

$$S = e_y^T P_y e_y + e_x^T P_x e_x.$$

According to the ideas of M-methods (Huber [8]), the residual functions, $e_y^T P_y e_y$, $e_x^T P_x e_x$ are replaced by the slow-growing residual functions $\sum_{i=1}^{n} r_i \rho(e_{y_i})$, $\sum_{j=1}^{n} s_j \rho(e_{x_j})$, respectively, to reduce or eliminate the influence of gross errors on the estimated parameters. The resulted estimation criterion is as follows:

$$\Omega = \sum_{i=1}^{n} r_i \rho(e_{y_i}) + \sum_{j=1}^{n} s_j \rho(e_{x_j}) = \min.$$

The corresponding Lagrange target function satisfies that

$$\Phi(e_y, e_x, \lambda, \xi) = \sum_{i=1}^{n} r_i \rho(e_{y_i}) + \sum_{j=1}^{n} s_j \rho(e_{x_j}) + \lambda^T \left(y - \beta_1 \cdot 1_n - \beta_2 \cdot (x - e_x) - e_y \right), \tag{4}$$

where λ is its Lagrange multiplier vector.

Then the Lagrange necessary conditions are obtained as:

$$\left.\frac{\partial \Phi}{\partial e_y}\right|_{\tilde{e}_y, \tilde{e}_x, \hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2} = \begin{bmatrix} r_1 \rho'(\tilde{e}_{y1}) \\ r_2 \rho'(\tilde{e}_{y2}) \\ \vdots \\ r_m \rho'(e_{yn}) \end{bmatrix} - \hat{\lambda} = 0, \tag{5}$$

$$\left.\frac{\partial \Phi}{\partial e_x}\right|_{\tilde{e}_y, \tilde{e}_x, \hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2} = \begin{bmatrix} s_1 \rho'(\tilde{e}_{x_1}) \\ s_2 \rho'(\tilde{e}_{x_2}) \\ \vdots \\ s_n \rho'(\tilde{e}_{x_n}) \end{bmatrix} + \hat{\beta}_2 \cdot \hat{\lambda} = 0, \tag{6}$$

$$\left.\frac{\partial \Phi}{\partial \lambda}\right|_{\tilde{e}_y, \tilde{e}_x, \hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2} = y - \hat{\beta}_1 \cdot 1_n - \hat{\beta}_2 \cdot (x - \tilde{e}_x) - \tilde{e}_y = 0, \tag{7}$$

$$\left.\frac{\partial \Phi}{\partial \beta_1}\right|_{\tilde{e}_y, \tilde{e}_x, \hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2} = -1_n{}^T \cdot \hat{\lambda} = 0, \tag{8}$$

$$\left.\frac{\partial \Phi}{\partial \beta_2}\right|_{\tilde{e}_y, \tilde{e}_x, \hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2} = \tilde{e}_x^T \hat{\lambda} - x^T \hat{\lambda} = 0. \tag{9}$$

Let

$$u(x) = \begin{cases} \dfrac{\rho'(x)}{x}, & x \neq 0, \\ 1, & x = 0, \end{cases}$$

and

$$w(e_y) = \begin{bmatrix} u(e_{y_1}) & & & \\ & u(e_{y_2}) & & \\ & & \ddots & \\ & & & u(e_{y_n}) \end{bmatrix}, \; w(e_x) = \begin{bmatrix} u(e_{x_2}) & & & \\ & u(e_{x_2}) & & \\ & & \ddots & \\ & & & u(e_{x_n}) \end{bmatrix}. \tag{10}$$

Equations (5) and (6) can be written, respectively, as

$$\left(P_y w(\tilde{e}_y)\right)\tilde{e}_y - \hat{\lambda} = 0, \tag{11}$$

$$(P_x w(\tilde{e}_x))\tilde{e}_x + \hat{\beta}_2 \cdot \hat{\lambda} = 0. \tag{12}$$

Now, $\tilde{e}_y$ and $\tilde{e}_x$ can be expressed in terms of $\hat{\lambda}$ by using (11) and (12). This leads to

$$\tilde{e}_y = \left(P_y w(\tilde{e}_y)\right)^{-1}\hat{\lambda},$$

$$\tilde{e}_x = -\hat{\beta}_2 (P_x w(\tilde{e}_x))^{-1}\hat{\lambda},$$

and after inserting this into (7), we have

$$\hat{\lambda} = \left[\left(P_y w(\tilde{e}_y)\right)^{-1} + \hat{\beta}_2^2 (P_x w(\tilde{e}_x))^{-1}\right]^{-1}\left(y - \hat{\beta}_1 \cdot 1_n - \hat{\beta}_2 \cdot x\right). \tag{13}$$

Let

$$Q_1 = \left(P_y w(\tilde{e}_y)\right)^{-1} + \hat{\beta}_2^2 (P_x w(\tilde{e}_x))^{-1},$$

and we readily obtain

$$\hat{\lambda} = Q_1^{-1}\left(y - \hat{\beta}_1 \cdot 1_n - \hat{\beta}_2 \cdot x\right). \tag{14}$$

Inserting (14) into (8), we get

$$\hat{\beta}_1 = \left(1_n{}^T Q_1^{-1} 1_n\right)^{-1} 1_n{}^T Q_1^{-1}\left(y - \hat{\beta}_2 \cdot x\right). \tag{15}$$

Let

$$Q_2 = x^T Q_1^{-1}\left(I_n - 1_n\left(1_n{}^T Q_1^{-1} 1_n\right)^{-1} 1_n{}^T Q_1^{-1}\right),$$

and

$$\tilde{e}_x{}^T \hat{\lambda} = \hat{\lambda}^T \tilde{e}_x.$$

Then we obtain from (9) that

$$\hat{\beta}_2 = (Q_2 x)^{-1}\left(Q_2 y - \hat{\lambda}^T \tilde{e}_x\right). \tag{16}$$

Inserting (16) into (15), the closed-form expression of the estimated parameter vector $\hat{\beta}_1$ and $\hat{\beta}_2$ can be derived as

$$\begin{cases} \hat{\beta}_1 = \left(1_n{}^T Q_1^{-1} 1_n\right)^{-1} 1_n{}^T Q_1^{-1}\left((I_n - (Q_2 x)^{-1} x Q_2)y + (Q_2 x)^{-1} x\hat{\lambda}^T \tilde{e}_x\right), \\ \hat{\beta}_2 = (Q_2 x)^{-1}\left(Q_2 y - \hat{\lambda}^T \tilde{e}_x\right). \end{cases} \tag{17}$$

Different robust estimators can be obtained by using different M-estimation functions.

3. Conclusion

Considering the random errors and gross errors may exist in both of the response variable and independent variable, the robust estimators are constructed in this paper; thus, this model in theoretically more rigorous than general robust estimation methods, which only considers the errors in the response variable. According to the idea of M-methods, by replacing the sum of residual squares of the response variable and independent variable with the slow-growing residual functions, a closed form solutions are obtained.

Acknowledgement

The research is supported by the Fundamental Research Funds for Central Universities (No.2652015193), Undergraduate Teaching-Reform Project (No.JGYB201529), Graduate Teaching-Reform Project of China of University of Geosciences (Beijing), NSFC (41374016, 41104025 and 41330634), and SKLGED2013-4-8-E.

References

[1] R. J. Adcock, Note on the method of least squares, Analyst 4(6) (1977), 183-184.

[2] C. L. Cheng and J. W. Van Ness, Statistical Regression with Measurement Error, Arnold and New York: Oxford University Press, London, 1999.

[3] C. L. Cheng and J. Riu, On estimating linear relationships when both variables are subject to heteroscedastic measurement errors, Technometrics 48(4) (2006), 511-519.

[4] W. A. Fuller, Measurement Error Models, John Wiley & Sons, New York, 2009.

[5] J. W. Gillard and T. C. Iles, Method of moments estimation in linear regression with errors in both variables, Cardiff University, School of Mathematics, Technical Paper, 2005.

[6] G. H. Golub and C. F. Van Loan, An analysis of the total least squares problem, SIAM Journal on Numerical Analysis 17(6) (1980), 883-893.

[7] P. J. Huber, Robust regression: Asymptotics, conjectures and Monte Carlo, The Annals of Statistics (1973), 799-821.

[8] P. J. Huber, Robust Statistics, Springer, Berlin Heidelberg, 2011.

[9] C. H. Kummel, Reduction of observed equations which contain more than one observed quantity, The Analyst (6) (1979), 97-105.

[10] V. Mahboub, A. R. Amiri-Simkooei and M. A. Sharifi, Iteratively reweighted total least squares: A robust estimation in errors in-variables models, Survey Review 45(329) (2013), 92-99.

[11] B. Schaffrin, I. Lee, Y. Felus and Y. Choi, Total least-squares for geodetic straight-line and plane adjustment, Bollettino Di Geodesia E Scienze Affini 65(3) (2006), 141-168.

[12] B. Schaffrin and A. Wieser, On weighted total least-squares adjustment for linear regression, Journal of Geodesy 82(7) (2008), 415-421.

[13] M. Y. Wong, Likelihood estimation of a simple linear regression model when both variables have error, Biometrika 76(1) (1989), 141-148.

■