Minimizing Oblique Errors for Robust Estimating

DIARMUID O'DRISCOLL, DONALD E. RAMIREZ,
AND REBECCA SCHMITZ

Abstract. The slope of the best fit line from minimizing the sum of the squared oblique errors is shown to be the root of a polynomial of degree four. We introduce a median estimator for the slope and, using a case study, we show that the median estimator is robust.

1. Introduction

With ordinary least squares (OLS) regression, we have data

$$\{(x_1, Y_1|X = x_1), \ldots, (x_n, Y_n|X = x_n)\}$$

and we minimize the sum of the squared *vertical errors* to find the *best-fit* line $y = h(x) = \beta_0 + \beta_1 x$. With OLS it is assumed that the independent or causal variable is measured without error.

J. L. Gill [2] states that "some regression prediction or estimation must be made in a direction opposite to the natural causality of one variable by another." This is found from the inverse function $h^{-1}(y_0) = x_0 = y_0/\beta_1 - \beta_0/\beta_1$. He adds "Geometric mean regression could be more valid than either direct or inverse regression if both variables are subject to substantial measurement error."

For inverse prediction we will want both $h(x)$ and $h^{-1}(y)$ to model the data. To accomplish this, we try to determine a fit so that the squared vertical and the squared horizontal errors will both be small. The vertical errors are the squared distances from (x, y) to $(x, h(x))$ and the horizontal errors are the squared distances from (x, y) to $(h^{-1}(y), y)$. As a compromise, we will consider the errors at the median or midpoint to the predicted vertical and predicted horizontal values. All of the estimated regression models we consider (including the geometric mean and perpendicular methods) are contained in the parametrization (with $0 \leq \lambda \leq 1$) of the line from $(x, h(x))$

to $(h^{-1}(y), y)$. For the squared vertical errors, set $\lambda = 1$ and correspondingly, for the horizontal errors, set $\lambda = 0$. Our Maple codes and the data set for our case study can be found here:

`people.virginia.edu/∼der/pdf/oblique_errors`

Our paper first introduces the Oblique Error Method in Section 2. In Section 3, we show how the Geometric Mean and Perpendicular Methods are included in our parametrization. In Section 4, we include a weighted regression procedure and Section 5 contains a small case study showing the robustness of the proposed median slope estimator.

2. Minimizing Squared Oblique Errors

From the data point (x_i, y_i) to the fitted line $y = h(x) = \beta_0 + \beta_1 x$ the vertical length is $a_i = |y_i - \beta_0 - \beta_1 x_i|$, the horizontal length is $b_i = |x_i - (y_i - \beta_0)/\beta_1| = |(\beta_1 x_i - y_i + \beta_0)/\beta_1| = |a_i/\beta_1|$ and the perpendicular length is $h_i = a_i/\sqrt{1 + \beta_i^2}$. With standard notation,

$$S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2, S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2, S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

with the correlation $\rho = S_{xy}/\sqrt{S_{xx}S_{yy}}$. A basic fact is $-1 \le \rho \le 1$ or equivalently $0 \le S_{xy}^2 \le S_{xx}S_{yy}$.

For the oblique length from (x_i, y_i) to $(h^{-1}(y_i) + \lambda(x_i - h^{-1}(y_i)), y_i + \lambda(h(x_i) - y_i))$, the horizontal length is $(1-\lambda)b_i = (1-\lambda)a_i/\beta_1$ and the vertical length is λa_i. Since $SSE_h(\beta_0, \beta_1, \lambda) = \left(\sum_{i=1}^n a_i^2\right)/\beta_1^2$ and $SSE_v(\beta_0, \beta_1, \lambda) = \sum_{i=1}^n a_i^2$, we have

$$\begin{aligned} SSE_o(\beta_0, \beta_1, \lambda) &= (1-\lambda)^2 SSE_h + \lambda^2 SSE_v \\ &= \sum_{i=1}^{n}\left\{\frac{(1-\lambda)^2 a_i^2}{\beta_1^2} + \lambda^2 a_i^2\right\} = \frac{(1-\lambda)^2 + \lambda^2\beta_1^2}{\beta_1^2}\sum_{i=1}^{n} a_i^2. \end{aligned}$$

Setting $\partial SSE_o/\partial\beta_0 = 0$, then $\beta_0 = \overline{y} - \beta_1\overline{x}$ and

$$\begin{aligned} \sum_{i=1}^{n} a_i^2 &= \sum_{i=1}^{n}\left\{(y_i - \overline{y}) - \beta_1(x_i - \overline{x})\right\}^2 \\ &= S_{yy} - 2\beta_1 S_{xy} + \beta_1^2 S_{xx}. \end{aligned}$$

Hence

$$SSE_o = ((1-\lambda)^2\beta_1^{-2} + \lambda^2)\left(S_{yy} - 2\beta_1 S_{xy} + \beta_1^2 S_{xx}\right) \qquad (1)$$

with

$$\frac{\partial SSE_o}{\partial \beta_1} = -2(1-\lambda)^2\beta_1^{-3}S_{yy}+2(1-\lambda)^2\beta_1^{-2}S_{xy}-2\lambda^2 S_{xy}+2\lambda^2\beta_1 S_{xx}.$$

Thus the oblique estimator is a root of the fourth degree polynomial in β_1, namely

$$P_4(\beta_1) = \lambda^2\sqrt{\frac{S_{xx}}{S_{yy}}}\beta_1^4 - \lambda^2\rho\beta_1^3 + (1-\lambda)^2\rho\beta_1 - (1-\lambda)^2\sqrt{\frac{S_{yy}}{S_{xx}}}. \quad (2)$$

We claim that $P_4(\beta_1)$ has exactly two real roots, one positive and one negative. By inspection, since the leading coefficient of $P_4(\beta_1)$ is positive and the constant coefficient is negative, $P_4(\beta_1)$ necessarily has at least one positive and one negative root. That these are the only real roots will be important in establishing the global minimum value for SSE_o.

The Complete Discrimination System $\{D_1,\ldots,D_n\}$ of Yang [4] is a set of explicit expressions that determine the number (and multiplicity) of roots of a polynomial. In the case of a fourth degree polynomial, the polynomial has exactly two real roots, each with multiplicity one, provided $D_4 < 0$; where $D_4 = 256a_0^3a_4^3+...+144a_0^2a_2a_4a_3^2$. The expression for D_4 has 16 terms involving the five coefficients $\{a_0,\ldots,a_4\}$ of the polynomial and it is of order 6.

For the polynomial $P_4(\beta_1)$ (with some manipulations),

$$\begin{aligned} D_4 &= \lambda^6(1-\lambda)^6(-256+192\rho^2+6\rho^4+4\rho^6) \\ &\quad -27\lambda^4(1-\lambda)^4\rho^4\left(\frac{S_{xx}}{S_{yy}}(1-\lambda)^4+\lambda^4\frac{S_{yy}}{S_{xx}}\right). \end{aligned}$$

Since $|\rho| \le 1$, it follows that $D_4 < 0$. And thus $P_4(\beta_1)$ has exactly one positive and one negative root.

Evaluating $\partial SSE_o/\partial\beta_1$ at $\beta_1 = S_{xy}/S_{xx}$ and using the inequality $0 \le S_{xy}^2 \le S_{xx}S_{yy}$ and the equality $S_{xx}S_{yy} - S_{xy}^2 = (1-\rho^2)S_{xx}S_{yy}$,

$$\begin{aligned} \frac{\partial SSE_o}{\partial \beta_1} &= \frac{-2(1-\lambda)^2}{\beta_1^2}\left\{\frac{S_{yy}}{S_{xy}/S_{xx}} - S_{xy}\right\} + 2\lambda^2\left\{-S_{xy}+\frac{S_{xy}}{S_{xx}}S_{xx}\right\} \\ &= \frac{-2(1-\lambda)^2}{\beta_1^2}\frac{1}{S_{xy}}S_{xx}S_{yy}(1-\rho^2) \end{aligned}$$

which has the sign of $-S_{xy}$. Similarly evaluating $\partial SSE_o/\partial\beta_1$ at $\beta_1 = S_{yy}/S_{xy}$

$$\frac{\partial SSE_o}{\partial\beta_1} = 2\lambda^2 \frac{1}{S_{xy}} S_{yy}S_{xx}(1-\rho^2)$$

which has the sign of S_{xy}.

We use the Intermediate Value Theorem to assert that (1) If $S_{xy} > 0$, then $0 < S_{xy}/S_{xx} \le \beta_1 \le S_{yy}/S_{xy}$; (2) If $S_{xy} < 0$, then $S_{yy}/S_{xy} \le \beta_1 \le S_{xy}/S_{xx} < 0$; and (3) If $S_{xy} = 0$, $\beta_1 = \pm\left(((1-\lambda)^2 S_{yy})/(\lambda^2 S_{xx})\right)^{1/4}$.

The Second Derivative Test assures that a root of $P_4(\beta_1)$ is a local minimum of SSE_o by

$$\frac{\partial^2 SSE_o}{\partial\beta_1^2} = \frac{6(1-\lambda)^2 S_{yy}}{\beta_1^4} - \frac{4(1-\lambda)^2 S_{xy}}{\beta_1^3} + 2\lambda^2 S_{xx}$$

$$= \frac{2(1-\lambda)^2}{\beta_1^4}\left[3S_{yy} - 2\beta_1 S_{xy}\right] + 2\lambda^2 S_{xx},$$

with $3S_{yy} - 2\beta_1 S_{xy} = 3S_{yy} - 2|\beta_1 S_{xy}| \ge 3S_{yy} - 2S_{yy} = S_{yy} > 0$.

Suppose $S_{xy} > 0$. Note from Equation (1) that $SSE_o(|\beta_1|) < SSE_o(-|\beta_1|)$. Let β_1^+ be the positive root of $P_4(\beta_1)$ and let β_1^- be the negative root of $P_4(\beta_1)$. Then $SSE_o(\beta_1^+) \le SSE_o(|\beta_1^-|) < SSE_o(\beta_1^-)$. This assures that the positive root gives the global minimum for $SSE_0(\beta_1)$. A similar result holds when $S_{xy} < 0$.

3. Minimizing Squared Perpendicular and Squared Geometric Mean Errors

The perpendicular error model dates back to Adcock [1] who introduced it as a procedure for fitting a straight line model to data with error measured in both the x and y directions.

For squared perpendicular errors we minimize $SSE_p(\beta_0, \beta_1) = \sum_{i=1}^n a_i^2/(1+\beta_1^2)$ with solutions $\beta_0^p = \overline{y} - \beta_1^p \overline{x}$ and

$$\beta_1^p = \frac{(S_{yy} - S_{xx}) \pm \sqrt{(S_{yy}-S_{xx})^2 + 4S_{xy}^2}}{2S_{xy}}, \tag{3}$$

(provided $S_{xy} \ne 0$).

Note with $S_{xy} \ne 0$ and $S_{xx} = S_{yy}$, then $\beta_1^p = \pm 1$ showing that under standardization this method is functionally independent of the correlation between x and y!

For squared geometric mean errors, we minimize $SSE_g(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(\sqrt{|a_i b_i|}\right)^2 = \sum_{i=1}^{n} a_i^2/|\beta_1|$ with solutions $\beta_0^g = \overline{y} - \beta_1^g \overline{x}$ and $\beta_1^g = \pm\sqrt{S_{yy}/S_{xx}}$. Note that β_1^g is always functionally independent of the correlation between x and y and also under standardization $b_1^g = \pm 1$ as in the perpendicular model.

The solutions to the above equations for both β_1^p and β_1^g are also roots of $P_4(\beta_1)$ for particular values of λ which can be seen from the geometry of the model. See [3] and [2] for applications of the perpendicular and geometric mean estimators.

4. Minimizing Squared Weighted Average Errors

If the user wishes to incorporate the effect of different variances in x and y, this can be achieved by using a weighed average of the squared vertical and squared horizontal errors with $(0 \le \alpha \le 1)$ and $SSE_w = \alpha SSE_v + (1-\alpha) SSE_h$. A typical value for α might be $\alpha = \sigma_y^2/(\sigma_x^2 + \sigma_y^2)$ to standardize the data. Recall from Section 2 that $SSE_o = \lambda^2 SSE_v + (1-\lambda)^2 SSE_h$. On setting $(1-\lambda)^2/\lambda^2 = (1-\alpha)/\alpha$, we get the quadratic equation $(2\alpha - 1)\lambda^2 - 2\alpha\lambda + \alpha = 0$, which has root

$$\lambda = \begin{cases} \dfrac{\alpha - \sqrt{\alpha(1-\alpha)}}{(2\alpha - 1)} & \alpha \neq \dfrac{1}{2} \\ \dfrac{1}{2} & \alpha = \dfrac{1}{2}. \end{cases} \tag{4}$$

5. Case Study

In this section, we introduce the median estimator β_1^m using $P_4(\beta_1)$ with $\lambda = 1/2$. Our small case study reveals the desirable robustness inherent in the median estimator. The data set is from [2] with $n = 40$. The case study shows that the perpendicular estimator is highly influenced by outliers in the data, with the vertical and horizontal estimators also being significantly influenced by outliers. The geometric mean estimator, as expected, is more robust; and our median estimator, introduced in this paper, being the most robust in this case study. For the Weighted Average procedure, $\alpha = S_{yy}/(S_{yy} + S_{xx}) = 0.671$ which from Equation 4 yields $\lambda = 0.588$.

The first table below gives the values for the slope β_1, y -intercept β_0, λ, and SSE. To study the effect of outliers, we pick a row from the data set and perturb the values by some factor.

The second table contains the basic values and, in addition, the square of the shifts in the slope and y-intercept caused by perturbing the x-data by a factor of 7.5 for the data point for case $k = 5$. Note that the median estimator has the smallest squared shift distance. The third table shows similar values by perturbing the y-data by a factor of 0.5 for case $k = 5$. Note that the perpendicular model has been greatly influenced by this one outlier.

	Vert	Horiz	Perp	Geom	Median	Wt Avg
β_1	1.28	1.59	1.48	1.43	1.38	1.35
β_0	136	104	115	121	126	130
λ	1.00	0.00	0.312	0.412	0.500	0.588
SSE	12565	6163	4330	4494	4908	5581

Table 1. Gill Data for Vertical (Vert), Horizontal (Horiz), Perpendicular (Perp), Geometric Mean (Geom), Median and Weighted Average (Wt Avg) Procedures

	Vert	Horiz	Perp	Geom	Median
β_1	0.0937	2.33	0.118	0.467	0.654
β_0	259	−4.39	256	215	193
SSE	62007	284364	61327	95987	129360
$(\beta_1^* - \beta_1)^2$	1.41	0.541	1.87	0.923	0.531
$(\beta_0^* - \beta_0)^2$	15040	11723	19855	8818	4506

Table 2. Gill Data perturbed with $x^*[5] = 7.5\ x[5]$

	Vert	Horiz	Perp	Geom	Median
β_1	0.875	1.99	1.51	1.32	1.23
β_0	174	57.2	107	127	137
SSE	30977	17770	13339	13841	14521
$(\beta_1^* - \beta_1)^2$	0.165	0.161	0.000717	0.0116	0.0228
$(\beta_0^* - \beta_0)^2$	1410	4875	3446	2789	2312

Table 3. Gill Data perturbed with $y^*[4] = 0.5\ y[4]$

We replicated the above perturbation procedure for each of the $n = 40$ cases and record in Table 4 and Table 5 the average squared change in slope and the average squared change in the y-intercept denoted $\{E(\beta_1^* - \beta_1)^2, E(\beta_0^* - \beta_0)^2\}$ by perturbing the original x-data and y-data values by a factor of $\{7.5, 0.5\}$ respectively. Table 6 records the average squared changes where the data has been jointly perturbed for $(x[k], y[k])$ by the factors $\{7.5, 0.5\}$ respectively.

	Vert	Horiz	Perp	Geom	Median
$E(\beta_1^* - \beta_1)^2$	1.41	4.01	1.87	1.07	0.656
$E(\beta_0^* - \beta_0)^2$	14966	57192	19820	10358	5649

Table 4. Gill Data perturbed with $x^*[k] = 7.5\ x[k]$

	Vert	Horiz	Perp	Geom	Median
$E(\beta_1^* - \beta_1)^2$	0.0163	0.201	0.0968	0.0333	0.0143
$E(\beta_0^* - \beta_0)^2$	165	2509	1276	488	229

Table 5. Gill Data perturbed with $y^*[k] = 0.5\ y[k]$

	Vert	Horiz	Perp	Geom	Median
$E(\beta_1^* - \beta_1)^2$	1.90	20.6	2.59	0.975	0.487
$E(\beta_0^* - \beta_0)^2$	20175	258272	27880	8611	3364

Table 6. Gill Data perturbed with $\{x^*[k] = 7.5\ x[k], y^*[k] = 0.5\ y[k]\}$

The results in Table 4 with an outlier in the x-data show the sensitivity with the vertical, horizontal and perpendicular procedures. The results in Table 5 with an outlier in the y-data show the sensitivity with the horizontal and perpendicular procedures. Table 6, with (x, y) both perturbed, shows the robustness of the geometric and median procedures with the median estimators uniformly superior to the geometric estimators in this small case study. These preliminary results commend the method for further investigation.

References

[1] R. J. Adcock, A problem in least-squares, *The Analyst*, **5** (1878), 53–54.
[2] J. L. Gill, Biases in regression when prediction is inverse to causation, *American Society of Animal Science* **64** (1987), 594–600.
[3] L. Leng, T. Zhang, L. Kleinman, and W. Zhu, Ordinary least square regression, orthogonal regression, geometric mean regression and their applications in aerosol science, *Journal of Physics: Conference Series* **78** (2007), 012084–012088.
[4] L. Yang, Recent advances on determining the number of real roots of parametric polynomials, *J. Symbolic Computation* **28** (1999), 225–242.

Diarmuid O'Driscoll,
Department of Mathematics and Computer Studies,
Mary Immaculate College,
Limerick, Ireland,
diarmuid.odriscoll@mic.ul.ie

Donald E. Ramirez,
Department of Mathematics,
University of Virginia,
Charlottesville, VA 22904, USA
der@virginia.edu

Rebecca Schmitz,
ITCEP,
University of Minnesota,
Minneapolis, MN 55455, USA
schm2676@umn.edu

Received on 2 October 2008.