2013

Generalized Least-Squares Regressions I: Efficient Derivations

Nataniel Greene
CUNY Kingsborough Community College

Recommended Citation

N. Greene. "Generalized Least-Squares Regressions I: Efficient Derivations," in Proceedings of the 1st International Conference on Computational Science and Engineering (CSE'13), 2013, pp. 145-158.

Generalized Least-Squares Regressions I: Efficient Derivations

NATANIEL GREENE
Department of Mathematics and Computer Science
Kingsborough Community College, CUNY
2001 Oriental Boulevard, Brooklyn, NY 11235
UNITED STATES
ngreene.math@gmail.com

Abstract: Ordinary least-squares regression suffers from a fundamental lack of symmetry: the regression line of *y* given *x* and the regression line of *x* given *y* are not inverses of each other. Alternative symmetric regression methods have been developed to address this concern, notably: orthogonal regression and geometric mean regression. This paper presents in detail a variety of least-squares regression methods which may not have been known or fully explicated. The derivation of each method is made efficient through the use of Ehrenberg's formula for the ordinary least-squares error and through the extraction of a weight function *g(b)* which characterizes the regression. For every case of generalized least-squares, the error between the line and the data is shown to be a product of the weight function *g(b)* and Ehrenberg's error formula.

Key–Words: Least-squares, symmetric least-squares, weighted ordinary least-squares, orthogonal regression, geometric mean regression.

1 Ordinary and Alternative Least-Squares Regression

1.1 Overview

Ordinary least-squares regression dates back to the work of Legendre and Gauss and it is still the most commonly used method for finding a line of best fit through a given set of data points. The method minimizes the average vertical distance between the data and the line. Its exclusive mention in elementary statistics texts ensures that it is the only method that many users of regression know. Despite its widespread use, ordinary least-squares regression suffers from some fundamental limitations.

One major limitation is the lack of symmetry. Ordinary least-squares begins with a decision of the part of the user that one of the variables is the independent variable (call it x) and the other variable is the dependent variable (call it y). The independent variable is assumed to be perfectly known, whereas the dependent variable has error. In this case, one computes the ordinary least-squares regression line for y given x (OLS $y|x$). The equation is $y = a + bx$ where $a = \mu_y - b\mu_x$ and $b = \rho\frac{\sigma_y}{\sigma_x}$. While it is valid to use this regression line to predict the y-value which corresponds to a given x-value, it is not valid to use this line to predict the x-value which corresponds to a given y-value. When y is known exactly and x is subject to error one must compute the ordinary least-squares regression line of x given y (OLS $x|y$). Only from this regression line is it valid to predict the x-value corresponding to a given y-value. This problem of there being two regression lines for a given set of data can be restated as follows: the OLS $y|x$ line and the OLS $x|y$ line are not inverses of each other. Indeed, the inverse form of the OLS $x|y$ line is given by $y = a' + b'x$ where $a' = \mu_y - b'\mu_x$ and $b' = \frac{1}{\rho}\frac{\sigma_y}{\sigma_x}$.

A more general and robust assumption, is to assume that both x and y variables have error, and to allow for the possibility of functional interdependence between the variables. In this case one can construct symmetric least-squares regressions in which the errors in both the x and y variables and the line are minimized. Two known examples of symmetric regressions are orthogonal regression and geometric mean regression (GMR). Orthogonal regression minimizes the average perpendicular or diagonal error between the data points and the line and GMR minimizes the average area of the triangles formed by the data points and the line. In contrast, OLS $y|x$ minimizes the average vertical error between the data and the line and OLS $x|y$ minimizes the average horizontal error between the data and the line.

Another limitation of standard ordinary least-squares is the lack of transitivity: the OLS regression line for $z|x$ is not the composition of the OLS regression line for $z|y$ with the OLS line for $y|x$. Orthog-

ISBN: 978-960-474-319-3

onal regression, while symmetric, is not transitive. GMR is both symmetric and transitive which makes it an ideal choice for robust modeling. GMR regression is also scale invariant, which means that multiplying the x or y values of the data by a constant factor and regressing is equivalent to multiplying x or y variable of the original regression line by that same constant. The formula for the GMR regression line is also particularly simple: $y = a + bx$ where $a = \mu_y - b\mu_x$ and $b = \operatorname{sgn}\rho\frac{\sigma_y}{\sigma_x}$. All these issues are discussed in greater detail in Taagepera [8].

Despite the great advantages of GMR for robust modelling, the goal of this paper and the papers which follow is to consider other possible generalized least-squares regressions as well, and to develop a theory for deriving and classifying these methods.

This paper has several specific goals: to re-derive the known cases of generalized least-squares in a manner that is as brief and efficient as possible; to explore the derivation of several new cases of symmetric regression; to see in the process that every symmetric regression is actually a weighted ordinary least-squares regression; and to explore the derivation of several weighted ordinary least-squares regressions. The choice of specific cases worked out here is by necessity idiosyncratic and cannot be comprehensive. This problem is rectified in the paper to follow [3], where this work is generalized.

1.2 Ordinary Least-Squares y Given x: Vertical Regression

Ordinary least-squares regression seeks coefficients a and b which minimize the average square *vertical* deviation between the data $\left\{(x_i, y_i)_{i=1}^N\right\}$ and the line $y = a + bx$. The vertical deviation between a data point (x_i, y_i) and the line $y = a + bx$ is given in absolute value by $|y_i - (a + bx_i)|$. For convenience this is written as $|a + bx_i - y_i|$. The average square vertical deviation is therefore given by

$$E = \frac{1}{N}\sum_{i=1}^{N}(a + bx_i - y_i)^2 . \tag{1}$$

For notational simplicity the error is represented using E and it is clear that one can also write $E = \varepsilon^2$. To minimize this error function, the usual procedure is to set partial derivatives $E_a = 0$ and $E_b = 0$ and solve for a and b. The result is

$$a = \mu_y - b\mu_x \tag{2}$$

and

$$b = \rho\frac{\sigma_y}{\sigma_x} \tag{3}$$

where μ_x and μ_y are the mean values of x and y given by

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N}x_i \tag{4}$$

and

$$\mu_y = \frac{1}{N}\sum_{i=1}^{N}y_i \tag{5}$$

and σ_x and σ_y are standard deviations given by

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2} \tag{6}$$

and

$$\sigma_y = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \mu_y\right)^2}. \tag{7}$$

Finally, ρ is the coefficient of correlation given by

$$\rho = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)\left(y_i - \mu_y\right)/\left(\sigma_x\sigma_y\right). \tag{8}$$

Note that population parameter notation μ_x, μ_y, σ_x, σ_y, ρ and N is used throughout this work, but the same results hold when they are replaced with sample statistics $\overline{x}$, $\overline{y}$, s_x, s_y, r and n.

One must verify that these values for a and b minimize the error function $E(a, b)$. The test is to check that the Hessian matrix of second order partial derivatives

$$H = \begin{bmatrix} E_{aa} & E_{ba} \\ E_{ab} & E_{bb} \end{bmatrix} \tag{9}$$

is positive definite. This occurs when its first entry and its determinant are both positive. Since

$$H = 2\begin{bmatrix} 1 & \mu_x \\ \mu_x & \sigma_x^2 + \mu_x^2 \end{bmatrix} \tag{10}$$

and $\det H = 4\sigma_x^2$, these values for a and b minimize the error.

1.3 Alternative Derivation: Ehrenberg's Formula

There is an alternative derivation of ordinary least-squares regression described in Ehrenberg [1] which does not require calculus. One first expresses E in terms of μ_x, μ_y, σ_x, σ_y and ρ. Then one finds the values of a and b which minimizes the error using elementary algebra. This simple approach has apparently been overlooked by expositors of least-squares.

ISBN: 978-960-474-319-3

Theorem 1 *(Ehrenberg's formula)*

$$E = \sigma_y^2\left(1-\rho^2\right) + \left(b\sigma_x - \rho\sigma_y\right)^2 + \left(a + b\mu_x - \mu_y\right)^2 \tag{11}$$

Proof.

$$\begin{aligned} E &= \frac{1}{N}\sum_{i=1}^{N}\left(a + bx_i - y_i\right)^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}\Big(b\left(x_i - \mu_x\right) - \left(y_i - \mu_y\right) \\ &\quad + \left(a + b\mu_x - \mu_y\right)\Big)^2 \\ &= b^2\sigma_x^2 - 2b\rho\sigma_x\sigma_y + \sigma_y^2 + \left(a + b\mu_x - \mu_y\right)^2 \\ &= \sigma_y^2\left(1-\rho^2\right) + \left(b\sigma_x - \rho\sigma_y\right)^2 \\ &\quad + \left(a + b\mu_x - \mu_y\right)^2 \end{aligned}$$

■

The minimizing values for a and b cause the last two terms to equal zero. The error is then governed by $\sigma_y^2\left(1-\rho^2\right)$. Recall that quantity $1-\rho^2$ measures the scatter of the data cloud away from the line of best fit. The quantity ρ^2 is called the coefficient of determination and it measures the strength of the linear relationship. When $\rho^2 = 1$, $1-\rho^2 = 0$ and there is no scatter because all the data fall along the line $y = a + bx$. When $\rho^2 = 0$, $1-\rho^2 = 1$, the data are uncorrelated and there is maximum scatter. In this case $b = 0$ and the regression line is given by $y = \mu_y$.

In the (a, b) plane the level curves of E are concentric ellipses with axes rotated and translated away from the origin and the error function itself is an elliptical paraboloid in (a, b, E) space.

Ehrenberg's formula will play a key role throughout this paper, greatly simplifying the derivations of all subsequent regression formulas.

1.4 Ordinary Least-Squares x Given y: Horizontal Regression

In horizontal regression coefficients a and b are sought which minimize the average square *horizontal* deviation between the data $\left\{(x_i, y_i)_{i=1}^N\right\}$ and the line $y = a + bx$, which can be written equivalently as $x = \frac{1}{b}y - \frac{a}{b}$. The horizontal deviation between a data point (x_i, y_i) and the line is given in absolute value by $\left|x_i - \left(\frac{1}{b}y_i - \frac{a}{b}\right)\right|$ or $\left|\frac{a}{b} + x_i - \frac{1}{b}y_i\right|$. The average horizontal deviation is therefore given by

$$E = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{a}{b} + x_i - \frac{1}{b}y_i\right)^2 \tag{12}$$

One could minimize this error function directly. However, it is more fruitful to utilize the relation

$$\frac{a}{b} + x_i - \frac{1}{b}y_i = \frac{1}{b}\left(a + bx_i - y_i\right) \tag{13}$$

which is algebraically straightforward. Geometrically, it means that the vertical deviation between a data value and line and the horizontal deviation between a data value and the line are constrained by the slope of the line according to the relation

$$b = \frac{\text{Vertical deviation}}{\text{Horizontal deviation}}. \tag{14}$$

This observation is used in Woolley's derivation of geometric mean regression [10]. The relation is used repeatedly in this paper to simplify every derivation because it always allows for the extraction of a weight function from the error expression. The result is an equivalent least-squares error function

$$E = \left(\frac{1}{b^2}\right)\cdot\frac{1}{N}\sum_{i=1}^{N}\left(a + bx_i - y_i\right)^2 \tag{15}$$

or

$$\begin{aligned} E &= g(b)\Big(\sigma_y^2\left(1-\rho^2\right) + \left(b\sigma_x - \rho\sigma_y\right)^2 \\ &\quad + \left(a + b\mu_x - \mu_y\right)^2\Big) \end{aligned} \tag{16}$$

where the weight function is

$$g(b) = \frac{1}{b^2}. \tag{17}$$

Setting $E_a = 0$ and $E_b = 0$ and solving for a and b yields the desired expressions. The simplest way to do this in practice is to first substitute

$$a = \mu_y - b\mu_x \tag{18}$$

and eliminate a from the error function. Then solve for b in

$$\frac{d}{db}\left\{\frac{1}{b^2}\left(\sigma_y^2\left(1-\rho^2\right) + \left(b\sigma_x - \rho\sigma_y\right)^2\right)\right\} = 0. \tag{19}$$

The result is

$$b = \frac{1}{\rho}\frac{\sigma_y}{\sigma_x}. \tag{20}$$

The Hessian matrix

$$H = \frac{2}{b^2}\begin{bmatrix} 1 & \mu_x \\ \mu_x & \frac{\rho\sigma_x\sigma_y}{b} + \mu_x^2 \end{bmatrix} \tag{21}$$

is positive definite since

$$\det H = \frac{4}{b^4}\sigma_x\sigma_y\left(\frac{\rho}{b}\right) \tag{22}$$

and b has the same sign as ρ. In fact, substitute for b and obtain

$$\det H = 4\frac{\sigma_x^6\rho^6}{\sigma_y^4} \tag{23}$$

Therefore these values for a and b minimize the error.

1.5 Orthogonal Least-Squares: Diagonal Regression

An alternative known regression method which takes into account both x and y deviations implicitly is called orthogonal regression. Other names include perpendicular regression or Deming regression [4,5]. In this method one finds a and b which minimize the average square diagonal deviation between the data and the line. Recall that the distance d_i between a point (x_i, y_i) and the line $y = a + bx$ is given by

$$d_i = \frac{|a + bx_i - y_i|}{\sqrt{1+b^2}}. \tag{24}$$

Thus the average square diagonal deviation is given by

$$E = \frac{1}{1+b^2}\cdot\frac{1}{N}\sum_{i=1}^{N}(a+bx_i-y_i)^2 \tag{25}$$

or

$$\begin{aligned} E &= g(b)\left(\sigma_y^2\left(1-\rho^2\right)+(b\sigma_x-\rho\sigma_y)^2\right.\\ &\quad \left.+\left(a+b\mu_x-\mu_y\right)^2\right) \end{aligned} \tag{26}$$

where the weight function is

$$g(b) = \frac{1}{1+b^2}. \tag{27}$$

Set $E_a = 0$ and $E_b = 0$ and solve for a and b to obtain the desired values of a and b. The simplest way to do this is to again substitute $a = \mu_y - b\mu_x$ into the error function and then solve

$$\frac{d}{db}\left\{\frac{1}{1+b^2}\left(\sigma_y^2\left(1-\rho^2\right)+(b\sigma_x-\rho\sigma_y)^2\right)\right\} = 0 \tag{28}$$

for b. The result is a quadratic equation in b:

$$\rho\sigma_x\sigma_y b^2 + \left(\sigma_x^2 - \sigma_y^2\right)b - \rho\sigma_x\sigma_y = 0. \tag{29}$$

Solving for b yields two explicit solutions when $\rho \neq 0$:

$$b_\pm = \frac{\left(\sigma_y^2-\sigma_x^2\right) \pm \sqrt{\left(\sigma_x^2-\sigma_y^2\right)^2 + 4\rho^2\sigma_x^2\sigma_y^2}}{2\rho\sigma_x\sigma_y}. \tag{30}$$

Since $b_+ b_- = -1$, one value of the slope is always positive and one value is always negative. The Hessian matrix

$$H = \frac{2}{1+b^2}\begin{bmatrix} 1 & \mu_x \\ \mu_x & \frac{\rho\sigma_x\sigma_y}{b}+\mu_x^2 \end{bmatrix} \tag{31}$$

is positive definite when

$$\det H = \frac{4\sigma_x\sigma_y}{\left(1+b^2\right)^2}\left(\frac{\rho}{b}\right) > 0 \tag{32}$$

which occurs when b has the same sign as ρ. Therefore in order to minimize the error function, always choose the solution with the same sign as ρ.

When $\rho = 0$ it is natural to define $b = 0$, since $\lim_{\rho\to 0} b_\pm = 0$. This is also consistent with setting $\rho = 0$ in the quadratic equation.

For any right triangle with sides A, B and C, there is the Reciprocal Pythagorean Theorem [2,6] which is not well-known. It states that

$$\frac{1}{A^2} + \frac{1}{B^2} = \frac{1}{L^2} \tag{33}$$

where L is the altitude to the hypotenuse C. This is equivalent to the statement that the altitude squared is half the harmonic mean of the squares of the two shortest sides. Symbolically,

$$L^2 = \frac{1}{2}\mathbf{H}\left(A^2, B^2\right) \tag{34}$$

where $\mathbf{H}(x,y) = \frac{2xy}{x+y}$ is the harmonic mean of x and y. The above formula for the distance between a point and a line is now derived using this result. Observe that

$$\begin{aligned} d_i^2 &= \frac{1}{2}\mathbf{H}\left((a+bx_i-y_i)^2, \left(\frac{a}{b}+x_i-\frac{1}{b}y_i\right)^2\right) \\ &= \frac{(a+bx_i-y_i)^2\left(\frac{a}{b}+x_i-\frac{1}{b}y_i\right)^2}{(a+bx_i-y_i)^2+\left(\frac{a}{b}+x_i-\frac{1}{b}y_i\right)^2}. \end{aligned} \tag{35}$$

Replacing $\frac{a}{b} + x_i - \frac{1}{b}y_i$ with $\frac{1}{b}(a + bx_i - y_i)$ yields the standard expression $d_i^2 = \frac{(a+bx_i-y_i)^2}{1+b^2}$. In this way, orthogonal least-squares can be understood as minimizing the average harmonic mean of the square deviations in x and y. Although the terminology is not used, orthogonal regression could be rightfully called harmonic mean regression (HMR) in parallel to the method of geometric mean regression (GMR), which is discussed next.

ISBN: 978-960-474-319-3

1.6 Geometric Mean Regression: Least Absolute Areas

The method of geometric mean regression (GMR) [4,7,8,10], is arguably the most robust of all symmetric least-squares regressions in that it has the simplest formulas for the coefficients a and b, the regression remains valid for changes in units in the x and y variables, one can compose two or more regression lines and obtain a valid result. The method minimizes the average absolute product of the deviations in x and y. This is the same as the geometric mean of the square deviations in x and y. Observe that

$$\begin{aligned} d_i^2 &= \mathbf{G}\left((a+bx_i-y_i)^2, \left(\frac{a}{b}+x_i-\frac{1}{b}y_i\right)^2\right) \\ &= \sqrt{(a+bx_i-y_i)^2\left(\frac{a}{b}+x_i-\frac{1}{b}y_i\right)^2} \\ &= \left|(a+bx_i-y_i)\left(\frac{a}{b}+x_i-\frac{1}{b}y_i\right)\right| \end{aligned} \tag{36}$$

where $\mathbf{G}(x,y) = \sqrt{xy}$ is the geometric mean of x and y. Write

$$E = \frac{1}{N}\sum_{i=1}^{N}\left|(a+bx_i-y_i)\left(\frac{a}{b}+x_i-\frac{1}{b}y_i\right)\right|. \tag{37}$$

Replace $\frac{a}{b}+x_i-\frac{1}{b}y_i$ with $\frac{1}{b}(a+bx_i-y_i)$ and obtain a least-squares error function

$$E = \frac{1}{|b|}\cdot\frac{1}{N}\sum_{i=1}^{N}(a+bx_i-y_i)^2 \tag{38}$$

or

$$\begin{aligned} E &= g(b)\left(\sigma_y^2\left(1-\rho^2\right)+(b\sigma_x-\rho\sigma_y)^2\right. \\ &\quad \left.+\left(a+b\mu_x-\mu_y\right)^2\right) \end{aligned} \tag{39}$$

where the weight function is

$$g(b) = \frac{1}{|b|}. \tag{40}$$

Substitute $a = \mu_y - b\mu_x$ and solve

$$\frac{d}{db}\left\{\frac{1}{|b|}\left(\sigma_y^2\left(1-\rho^2\right)+(b\sigma_x-\rho\sigma_y)^2\right)\right\} = 0$$

for b. The result is

$$b = \pm\frac{\sigma_y}{\sigma_x} \tag{41}$$

The Hessian matrix

$$H = \frac{2}{|b|}\begin{bmatrix} 1 & \mu_x \\ \mu_x & \sigma_x^2+\mu_x^2 \end{bmatrix} \tag{42}$$

is positive definite because

$$\det H = \frac{4\sigma_x^2}{b^2} = \frac{4\sigma_x^4}{\sigma_y^2} \tag{43}$$

is always positive. Choose the sign of b to agree with the sign of ρ and write

$$b = \operatorname{sgn}\rho\frac{\sigma_y}{\sigma_x}. \tag{44}$$

It is natural to define $b = 0$ when $\rho = 0$. This is also consistent with the use of the sign function in the formula, since $\operatorname{sgn} 0 = 0$.

2 Alternative Symmetric Least-Squares Regressions

In this section alternative methods of least-squares regression are explored which may not have been known or fully explicated. The derivation for each method uses the same streamlined steps that were employed to derive the known methods. This is done with an eye toward the next paper in this series, where the derivation is generalized and all possible symmetric least-squares regressions are analyzed and classified.

2.1 Pythagorean Regression

An alternative method to orthogonal regression and GMR called Pythagorean least-squares is now investigated. The method is symmetric in that it minimizes the average square deviation in both x and y simultaneously. The error function is given by

$$E = \frac{1}{N}\sum_{i=1}^{N}\left((a+bx_i-y_i)^2+\left(\frac{a}{b}+x_i-\frac{1}{b}y_i\right)^2\right). \tag{45}$$

The term Pythagorean is appropriate because the method minimizes the average square hypotenuse of the triangles formed by the data points and the line. One can carry the usual minimization procedure on this expression or again replace $\frac{a}{b}+x_i-\frac{1}{b}y_i$ with $\frac{1}{b}(a+bx_i-y_i)$ and write

$$E = \left(1+\frac{1}{b^2}\right)\cdot\frac{1}{N}\sum_{i=1}^{N}(a+bx_i-y_i)^2 \tag{46}$$

ISBN: 978-960-474-319-3

or

$$E = g(b)\left(\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2 + \left(a+b\mu_x-\mu_y\right)^2\right) \quad (47)$$

where the weight function is

$$g(b) = 1+\frac{1}{b^2}. \quad (48)$$

As before, substitute $a = \mu_y - b\mu_x$ into the error function and then solve

$$\frac{d}{db}\left\{\left(1+\frac{1}{b^2}\right)\left(\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2\right)\right\} = 0 \quad (49)$$

for b. The result is a quartic equation in b:

$$b^4 - \rho\frac{\sigma_y}{\sigma_x}b^3 + \rho\frac{\sigma_y}{\sigma_x}b - \left(\frac{\sigma_y}{\sigma_x}\right)^2 = 0. \quad (50)$$

For practical purposes, this quartic polynomial must be solved numerically.

It is now proved that there is always a unique real solution to this equation with the same sign as ρ which minimizes the error function. The discriminant for a monic quartic polynomial $x^4+Bx^3+Cx^2+Dx+E$ is given by [9]

$$\begin{aligned}\Delta = {} & B^2C^2D^2 - 4B^2C^3E - 4B^3D^3 \\ & +18B^3CDE - 27B^4E^2 - 4C^3D^2 \\ & +16C^4E + 18BCD^3 - 80BC^2DE \\ & -6B^2D^2E + 144B^2CE^2 - 27D^4 \\ & +144CD^2E - 128C^2E^2 - 192BDE^2 \\ & +256E^3. \end{aligned} \quad (51)$$

When $\Delta < 0$ there are always two real solutions and two complex conjugate solutions. In our case

$$\begin{aligned}\Delta = {} & \left(4\rho^6+6\rho^4+192\rho^2-256\right)\left(\frac{\sigma_y}{\sigma_x}\right)^6 \\ & -27\rho^2\left(\left(\frac{\sigma_y}{\sigma_x}\right)^8+\left(\frac{\sigma_y}{\sigma_x}\right)^4\right)\end{aligned} \quad (52)$$

which is less than zero for $|\rho| \le 1$, the natural range of ρ. This implies that there are exactly two real solutions. Descartes' Rule of Signs can be used to further deduce the nature of the solutions.

Descartes' Rule of Signs states that for a polynomial $f(b)$ in standard form, the number of positive roots equals the number of changes in sign between consecutive nonzero coefficients, or it is less than this number by a multiple of two. The number of negative roots of f equals the number of changes in sign between consecutive nonzero coefficients of $f(-b)$, or it is less than this number by a multiple of two.

For $\rho > 0$, there are three changes in sign in $f(b)$, but since there are only two real solutions, the number of positive solutions must differ by two, and so there is one and only one positive real solution. The other real solution is negative. Similarly for $\rho < 0$, there is one change in sign in $f(-b)$, and so there is exactly one negative real solution. The other real solution is positive.

It is shown now that the real solution to the quartic equation is always greater than $\rho\frac{\sigma_y}{\sigma_x}$ for $\rho > 0$ and less than $\rho\frac{\sigma_y}{\sigma_x}$ for $\rho < 0$. Re-expand the polynomial in the variable $q = b - \rho\frac{\sigma_y}{\sigma_x}$ and call the resulting polynomial $h(q)$. The variable q measures the discrepancy between the Pythagorean slope and the ordinary least-squares slope. The result is the equivalent equation

$$\begin{aligned}0 = {} & q^4 + 3\rho\frac{\sigma_y}{\sigma_x}q^3 + 3\rho^2\left(\frac{\sigma_y}{\sigma_x}\right)^2q^2 \\ & + \left(\rho^3\left(\frac{\sigma_y}{\sigma_x}\right)^3 + \rho\frac{\sigma_y}{\sigma_x}\right)q \\ & -\frac{1}{2}\left(1-\rho^2\right)\left(\frac{\sigma_y}{\sigma_x}\right)^2.\end{aligned} \quad (53)$$

According to Descartes' Rule of Signs, when $\rho > 0$ there is one change in sign in $h(q)$and therefore one positive solution. When $\rho < 0$ there is one change in sign in $h(-q)$ and therefore one negative solution. Therefore there is one positive solution for b that is greater than $\rho\frac{\sigma_y}{\sigma_x}$ when ρ is positive and one negative solution for b that is less than $\rho\frac{\sigma_y}{\sigma_x}$ when ρ is negative.

The Hessian matrix

$$H = 2\left(1+\frac{1}{b^2}\right)\begin{bmatrix} 1 & \mu_x \\ \mu_x & \frac{\det H}{4\left(1+\frac{1}{b^2}\right)^2}+\mu_x^2\end{bmatrix} \quad (54)$$

is positive definite since

$$\begin{aligned}\det H = {} & 4\left(1+\frac{1}{b^2}\right)^2\sigma_x^2\left\{\left(\frac{\rho}{b}\right)\frac{\sigma_y}{\sigma_x} \right. \\ & \left. +\frac{4b\left(b-\rho\frac{\sigma_y}{\sigma_x}\right)}{b^2+1}\right\}\end{aligned} \quad (55)$$

is always positive for these values of b.

Therefore to perform Pythagorean regression, the user solves the quartic polynomial above numerically and chooses the one and only real solution for b with the same sign as ρ. This value for b is always greater in magnitude than b_{OLS} and it minimizes the error function along with the corresponding value for a.

2.2 Least Perimeter Squared Regression

Consider minimizing the average squared perimeter of the triangles formed by the data values and the regression line. The error function is

$$E=\frac{1}{N}\sum_{i=1}^{N}\left(|a+bx_i-y_i|+\left|\frac{a}{b}+x_i-\frac{1}{b}y_i\right|\right)^2 \tag{56}$$

and the resulting method is called least perimeter squared regression. Replace again $\frac{a}{b}+x_i-\frac{1}{b}y_i$ with $\frac{1}{b}(a+bx_i-y_i)$ and obtain

$$E=\left(1+\frac{1}{|b|}\right)^2\cdot\frac{1}{N}\sum_{i=1}^{N}(a+bx_i-y_i)^2 \tag{57}$$

or

$$\begin{aligned} E &= g(b)\left(\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2\right. \\ &\quad \left.+\left(a+b\mu_x-\mu_y\right)^2\right) \end{aligned} \tag{58}$$

where the weight function is

$$g(b)=\left(1+\frac{1}{|b|}\right)^2. \tag{59}$$

Substitute $a=\mu_y-b\mu_x$ into the error function and then solve

$$\frac{d}{db}\left\{\left(1+\frac{1}{|b|}\right)^2\left(\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2\right)\right\}=0 \tag{60}$$

for b. The result is a cubic equation in b

$$b^3-\rho\frac{\sigma_y}{\sigma_x}b^2+\operatorname{sgn}b\left(\rho\frac{\sigma_y}{\sigma_x}\right)b-\operatorname{sgn}b\left(\frac{\sigma_y}{\sigma_x}\right)^2=0. \tag{61}$$

Denote this polynomial by $f(b)$. It is now proved that there is always a unique real solution to this equation with the same sign as ρ. According to Descartes' Rule of Signs there are three changes in sign in $f(b)$ when $\rho>0$ implying either three positive solutions or one positive solution. There are also three changes in sign in $f(-b)$ when $\rho<0$ implying either three negative solutions or one negative solution.

However, the discriminant for a monic cubic polynomial x^3+Bx^2+Cx+D is given by [9]

$$\Delta=B^2C^2-4C^3-4B^3D+18BCD-2D^2. \tag{62}$$

When Δ is negative the cubic has one real solution and two complex conjugate solutions. When Δ is positive there are three real solutions. In our case

$$\begin{aligned} \Delta &= \left(\rho^4+18\left(\operatorname{sgn}b\right)\rho^2-27\right)\left(\frac{\sigma_y}{\sigma_x}\right)^4 \\ &\quad -4\left(\operatorname{sgn}b\right)\rho^3\left(\left(\frac{\sigma_y}{\sigma_x}\right)^5+\left(\frac{\sigma_y}{\sigma_x}\right)^3\right) \end{aligned} \tag{63}$$

and it is is guaranteed to be negative when $\operatorname{sgn}b=\operatorname{sgn}\rho$. Therefore there is always a unique real solution to the polynomial equation with the same sign as ρ.

The Hessian matrix

$$H=2\left(1+\frac{1}{|b|}\right)^2\begin{bmatrix}1 & \mu_x\\ \mu_x & \frac{\det H}{4(1+1/|b|)^4}+\mu_x^2\end{bmatrix} \tag{64}$$

is positive definite since

$$\begin{aligned} \det H &= 4\left(1+\frac{1}{|b|}\right)^4\sigma_x^2\left\{\left(\frac{\rho}{b}\right)\frac{\sigma_y}{\sigma_x}\right. \\ &\quad \left.+\frac{3\operatorname{sgn}b\left(b-\rho\frac{\sigma_y}{\sigma_x}\right)}{|b|+1}\right\}>0. \end{aligned} \tag{65}$$

The polynomial equation can be written unambiguously by substituting $\operatorname{sgn}b=\operatorname{sgn}\rho$ and $(\operatorname{sgn}b)\rho=|\rho|$ into the coefficient expressions, obtaining

$$b^3-\rho\frac{\sigma_y}{\sigma_x}b^2+|\rho|\frac{\sigma_y}{\sigma_x}b-\operatorname{sgn}\rho\left(\frac{\sigma_y}{\sigma_x}\right)^2=0. \tag{66}$$

In this form, the user solves the cubic polynomial for the one and only real solution. This solution is guaranteed to have the same sign as ρ, to be greater in magnitude than b_{OLS} and to minimize the error function. It is natural to define $b=0$ when $\rho=0$ which is consistent with the sign function which appears in the polynomial.

2.3 Squared Harmonic Mean Regression

It was seen that orthogonal least-squares can be thought of as minimizing the average harmonic mean of the square deviations in x and y. Consider instead minimizing the average square of the harmonic means of the absolute deviations in x and y. That is minimize

$$E=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{2\left|a+bx_i-y_i\right|\left|\frac{a}{b}+x_i-\frac{1}{b}y_i\right|}{\left|a+bx_i-y_i\right|+\left|\frac{a}{b}+x_i-\frac{1}{b}y_i\right|}\right)^2. \tag{67}$$

Replace $\frac{a}{b}+x_i-\frac{1}{b}y_i$ with $\frac{1}{b}(a+bx_i-y_i)$ and obtain

$$E=\frac{1}{(1+|b|)^2}\cdot\frac{1}{N}\sum_{i=1}^{N}(a+bx_i-y_i)^2 \tag{68}$$

ISBN: 978-960-474-319-3

or

$$E = g(b)\left(\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2 + \left(a+b\mu_x-\mu_y\right)^2\right) \quad (69)$$

where the weight function is

$$g(b) = \frac{1}{\left(1+|b|\right)^2}. \quad (70)$$

Substitute $a = \mu_y - b\mu_x$ into the error function and then solve

$$\frac{d}{db}\left\{\frac{\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2}{\left(1+|b|\right)^2}\right\} = 0 \quad (71)$$

for b. The result is the following formula for b:

$$b = \operatorname{sgn}\rho\frac{\sigma_y}{\sigma_x}\left(\frac{\frac{\sigma_y}{\sigma_x}+|\rho|}{|\rho|\frac{\sigma_y}{\sigma_x}+1}\right). \quad (72)$$

The Hessian matrix

$$H = \frac{2}{\left(1+|b|\right)^2}\begin{bmatrix} 1 & \mu_x \\ \mu_x & \frac{4\det H}{(1+|b|)^4}+\mu_x^2 \end{bmatrix} \quad (73)$$

is positive definite since

$$\det H = \frac{4\sigma_x^2\left(1+|\rho|\frac{\sigma_y}{\sigma_x}\right)}{\left(1+|b|\right)^5} \quad (74)$$

is always greater than zero. Therefore these formulas for a and b minimize the error.

3 Weighted Ordinary Least-Squares Regressions

It is apparent from the previous cases that a symmetric least-squares problem can be written as a weighted ordinary least-squares problem where a weight function multiplies the ordinary least-squares error. This section explores some weighted ordinary least-squares regression problems where the weight function is chosen from a related symmetric least absolute deviation regression problem. These problems are also referred to here as hybrid symmetric least-squares problems.

3.1 Hybrid Least Perimeter Regression

An alternative to least perimeter squared regression is to simply minimize the average perimeter of the triangles formed by the data values and the regression line The error function is then

$$D = \frac{1}{N}\sum_{i=1}^{N}\left(\left|a+bx_i-y_i\right|+\left|\frac{a}{b}+x_i-\frac{1}{b}y_i\right|\right). \quad (75)$$

Replace again $\frac{a}{b}+x_i-\frac{1}{b}y_i$ with $\frac{1}{b}\left(a+bx_i-y_i\right)$ and obtain

$$D = \left(1+\frac{1}{|b|}\right)\cdot\frac{1}{N}\sum_{i=1}^{N}\left|a+bx_i-y_i\right|. \quad (76)$$

This error function can certainly be minimized, however it is not a problem in least-squares but in least absolute deviation. Consider instead a hybrid symmetric least-squares problem with an error function that combines the weight function from the least absolute deviation problem and the ordinary least-squares error. The function is

$$E = \left(1+\frac{1}{|b|}\right)\cdot\frac{1}{N}\sum_{i=1}^{N}\left(a+bx_i-y_i\right)^2$$

or

$$E = g(b)\left(\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2 + \left(a+b\mu_x-\mu_y\right)^2\right) \quad (77)$$

where the weight function is

$$g(b) = 1+\frac{1}{|b|}. \quad (78)$$

Call the minimization of this error hybrid least-perimeter regression. This is a generalized least-squares problem which can be solved using the purely analytic methods used so far. Eliminate a as usual and solve

$$\frac{d}{db}\left\{\left(1+\frac{1}{|b|}\right)\left(\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2\right)\right\} = 0.$$

This results in a cubic equation (after choosing $\operatorname{sgn} b = \operatorname{sgn}\rho$)

$$b^3+\frac{\operatorname{sgn}\rho}{2}\left(1-2|\rho|\sigma_x\sigma_y\right)b^2-\frac{\operatorname{sgn}\rho}{2}\left(\frac{\sigma_y}{\sigma_x}\right)^2 = 0. \quad (79)$$

It is now proved using Descartes' Rule of Signs that there is always one positive real solution to this equation when ρ is positive and one negative real solution to this equation when ρ is negative. Let $f(b)$ denote the polynomial. When ρ is positive and $\left(1-2|\rho|\sigma_x\sigma_y\right)$ is positive, there is one change in sign and therefore one positive solution. When ρ

ISBN: 978-960-474-319-3

is positive and $(1-2|\rho|\sigma_x\sigma_y)$ is negative, there is again one change in sign and therefore one positive solution.

When ρ is negative and $(1-2|\rho|\sigma_x\sigma_y)$ is positive, there is one change in sign in $f(-b)$ and therefore one negative solution. When ρ is negative and $(1-2|\rho|\sigma_x\sigma_y)$ is negative, there is one change in sign in $f(-b)$ and therefore one negative solution.

The Hessian matrix

$$H = 2\left(1+\frac{1}{|b|}\right)\begin{bmatrix} 1 & \mu_x \\ \mu_x & \frac{\det H}{4(1+1/|b|)^2}+\mu_x^2 \end{bmatrix} \tag{80}$$

is positive definite since

$$\det H = 4\left(1+\frac{1}{|b|}\right)^2 \sigma_x^2 \left\{1+\frac{2\,\mathrm{sgn}\,b\left(b-\rho\frac{\sigma_y}{\sigma_x}\right)}{|b|+1}\right\} \tag{81}$$

is always greater than zero. Therefore these formulas for a and b minimize the error.

3.2 Hybrid Pythagorean Regression

This method reconsiders Pythagorean regression but instead of minimizing the average square hypotenuse of the triangles from by the data point and line, it minimizes the average hypotenuse. The corresponding error function is

$$D = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(a+bx_i-y_i)^2+\left(\frac{a}{b}+x_i-\frac{1}{b}y_i\right)^2} \tag{82}$$

Replace $\frac{a}{b}+x_i-\frac{1}{b}y_i$ with $\frac{1}{b}(a+bx_i-y_i)$ and obtain

$$D = \sqrt{1+\frac{1}{b^2}}\cdot\frac{1}{N}\sum_{i=1}^{N}|a+bx_i-y_i|\,. \tag{83}$$

This is another generalized least absolute deviation problem which can be solved using algorithms for least absolute deviation. The interest here however, is in generalized least-squares problems since they are amenable to analytic solutions. Consider instead a hybrid symmetric least-squares problem with error function

$$E = \sqrt{1+\frac{1}{b^2}}\cdot\frac{1}{N}\sum_{i=1}^{N}(a+bx_i-y_i)^2\,. \tag{84}$$

Replace with Ehrenberg's formula and write

$$\begin{aligned} E &= g(b)\Big(\sigma_y^2\left(1-\rho^2\right)+(b\sigma_x-\rho\sigma_y)^2 \\ &\quad + \left(a+b\mu_x-\mu_y\right)^2\Big) \end{aligned} \tag{85}$$

where the weight function is

$$g(b) = \sqrt{1+\frac{1}{b^2}}. \tag{86}$$

Call the minimization of this error function hybrid Pythagorean regression. Eliminate a as usual and solve

$$\frac{d}{db}\left[\sqrt{1+\frac{1}{b^2}}\left(\sigma_y^2\left(1-\rho^2\right)+(b\sigma_x-\rho\sigma_y)^2\right)\right]=0. \tag{87}$$

The result is a quartic polynomial

$$b^4-\rho\frac{\sigma_y}{\sigma_x}b^3+\frac{1}{2}b^2-\frac{1}{2}\left(\frac{\sigma_y}{\sigma_x}\right)^2=0. \tag{88}$$

Denote this polynomial by $f(b)$. According to Descartes' Rule of Signs there are three changes in sign in $f(b)$ when $\rho>0$ implying either three positive solutions or one positive solution. There are also three changes in sign in $f(-b)$ when $\rho<0$ implying either three negative solutions or one negative solution.

However, re-expand this polynomial in the variable $q=b-\rho\frac{\sigma_y}{\sigma_x}$ and call the resulting polynomial $h(q)$. The variable q measures the discrepancy between the hybrid Pythagorean slope and the ordinary least-squares slope. The result is the equivalent equation

$$\begin{aligned} 0 &= q^4+3\rho\frac{\sigma_y}{\sigma_x}q^3+\left(3\rho^2\left(\frac{\sigma_y}{\sigma_x}\right)^2+\frac{1}{2}\right)q^2 \\ &\quad +\left(\rho^3\left(\frac{\sigma_y}{\sigma_x}\right)^3+\rho\frac{\sigma_y}{\sigma_x}\right)q \\ &\quad -\frac{1}{2}\left(1-\rho^2\right)\left(\frac{\sigma_y}{\sigma_x}\right)^2. \end{aligned} \tag{89}$$

According to Descartes' Rule of Signs, when $\rho>0$ there is one change in sign in $h(q)$ and therefore one positive solution. When $\rho<0$ there is one change in sign in $h(-q)$ and therefore one negative solution. Therefore there is one positive solution for b that is greater than $\rho\frac{\sigma_y}{\sigma_x}$ when ρ is positive and one negative solution for b that is less than $\rho\frac{\sigma_y}{\sigma_x}$ when ρ is negative.

The Hessian matrix

$$H = \frac{2}{\sqrt{1+\frac{1}{b^2}}}\begin{bmatrix} 1 & \mu_x \\ \mu_x & \frac{\det H}{4\left(1+\frac{1}{b^2}\right)}+\mu_x^2 \end{bmatrix} \tag{90}$$

is guaranteed to be positive definite since

$$\det H = \frac{4\sigma_x^2}{\left(1+\frac{1}{b^2}\right)}\left\{1+\frac{3b\left(b-\rho\frac{\sigma_y}{\sigma_x}\right)}{b^2+1}\right\}>0 \tag{91}$$

for these values of b.

Therefore to perform hybrid Pythagorean regression, the user solves the quartic polynomial above numerically and chooses the unique real solution b with the same sign as ρ that is greater in magnitude than b_{OLS}. This value for b and the corresponding value for a minimize the error.

3.3 Hybrid Orthogonal Regression

Consider again orthogonal regression but instead of minimizing the average square diagonal distance between the data points and line, minimize the average diagonal distance. Since the distance d_i between a point (x_i, y_i) and the line $y = a + bx$ is given by

$$d_i = \frac{|a + bx_i - y_i|}{\sqrt{1+b^2}}.$$

Thus the average absolute diagonal deviation is given by

$$D = \frac{1}{\sqrt{1+b^2}} \cdot \frac{1}{N} \sum_{i=1}^{N} |a + bx_i - y_i| \,. \tag{92}$$

This is another generalized least absolute deviation problem which can be solved using algorithms for least absolute deviation. Consider here instead a hybrid error function

$$E = \frac{1}{\sqrt{1+b^2}} \cdot \frac{1}{N} \sum_{i=1}^{N} (a + bx_i - y_i)^2 \tag{93}$$

or

$$\begin{aligned} E &= g(b) \left(\sigma_y^2 \left(1 - \rho^2\right) + (b\sigma_x - \rho\sigma_y)^2 \right. \\ &\quad \left. + \left(a + b\mu_x - \mu_y\right)^2 \right) \end{aligned} \tag{94}$$

where the weight function is

$$g(b) = \frac{1}{\sqrt{1+b^2}}. \tag{95}$$

Call the minimization of this error function hybrid orthogonal regression. Eliminate a as usual and solve

$$\frac{d}{db} \left\{ \frac{1}{\sqrt{1+b^2}} \left(\sigma_y^2 \left(1 - \rho^2\right) + (b\sigma_x - \rho\sigma_y)^2 \right) \right\} = 0. \tag{96}$$

The result is a cubic polynomial equation

$$b^3 + \left(2 - \frac{\sigma_y}{\sigma_x}\right) b - 2\rho \frac{\sigma_y}{\sigma_x} = 0. \tag{97}$$

Let $f(b)$ denote the polynomial. According to Descartes' Rule of Signs, when $\rho > 0$ and $2 - \frac{\sigma_y}{\sigma_x} > 0$, $f(b)$ has one change in sign and there is one positive solution. When $\rho > 0$ and $2 - \frac{\sigma_y}{\sigma_x} < 0$, $f(b)$ has one change of sign and there is one positive solution. When $\rho < 0$ and $2 - \frac{\sigma_y}{\sigma_x} > 0$, $f(-b)$ has one change of sign and there is one negative solution. Finally when $\rho < 0$ and $2 - \frac{\sigma_y}{\sigma_x} < 0$, $f(-b)$ has one change of sign and there is one negative solution.

The Hessian matrix

$$H = \frac{2}{\sqrt{1+b^2}} \begin{bmatrix} 1 & \mu_x \\ \mu_x & \frac{\det H}{4(1+b^2)} + \mu_x^2 \end{bmatrix} \tag{98}$$

is positive definite since

$$\det H = \frac{4\sigma_x^2}{(1+b^2)} \left\{ \left(\frac{\rho}{b}\right) \frac{\sigma_y}{\sigma_x} + \frac{b\left(b - \rho\frac{\sigma_y}{\sigma_x}\right)}{1+b^2} \right\} > 0. \tag{99}$$

Therefore these formulas for a and b minimize the error.

3.4 Hybrid Harmonic Mean Regression

The simplest regression involving harmonic means of absolute deviations in x and y is to minimize

$$\begin{aligned} D &= \frac{1}{N} \sum_{i=1}^{N} \frac{2\,|a + bx_i - y_i| \left|\frac{a}{b} + x_i - \frac{1}{b}y_i\right|}{|a + bx_i - y_i| + \left|\frac{a}{b} + x_i - \frac{1}{b}y_i\right|} \\ &= \frac{1}{N} \sum_{i=1}^{N} \mathbf{H}\left(|a + bx_i - y_i| \,, \left|\frac{a}{b} + x_i - \frac{1}{b}y_i\right| \right) \end{aligned} \tag{100}$$

where $\mathbf{H}$ denotes the harmonic mean as before. Replacing $\frac{a}{b} + x_i - \frac{1}{b}y_i$ with $\frac{1}{b}(a + bx_i - y_i)$ yields

$$D = \frac{1}{(1+|b|)} \cdot \frac{1}{N} \sum_{i=1}^{N} |a + bx_i - y_i| \,. \tag{101}$$

Now consider here instead the hybrid error function

$$E = \frac{1}{(1+|b|)} \cdot \frac{1}{N} \sum_{i=1}^{N} (a + bx_i - y_i)^2 \tag{102}$$

or

$$\begin{aligned} E &= g(b) \left(\sigma_y^2 \left(1 - \rho^2\right) + (b\sigma_x - \rho\sigma_y)^2 \right. \\ &\quad \left. + \left(a + b\mu_x - \mu_y\right)^2 \right) \end{aligned} \tag{103}$$

where the weight function is

$$g(b) = \frac{1}{(1+|b|)}. \tag{104}$$

ISBN: 978-960-474-319-3

Call the minimization of this error function hybrid harmonic mean regression. Eliminate a as usual and solve

$$\frac{d}{db}\left\{\frac{1}{(1+|b|)}\left(\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2\right)\right\}=0. \tag{105}$$

The result is (after letting $\operatorname{sgn} b = \operatorname{sgn}\rho$)

$$b=\operatorname{sgn}\rho\left(-1+\sqrt{\left(\frac{\sigma_y}{\sigma_x}+|\rho|\right)^2+\left(1-\rho^2\right)}\right). \tag{106}$$

The Hessian matrix

$$H=\frac{2}{(1+|b|)}\left[\begin{array}{cc}1 & \mu_x\\ \mu_x & \sigma_x^2+\mu_x^2\end{array}\right] \tag{107}$$

is positive definite since

$$\det H=\frac{4\sigma_x^2}{(1+|b|)^2}>0. \tag{108}$$

Therefore these formulas for a and b minimize the error.

3.5 Exponentially-Weighted Ordinary Least-Squares Regression

An interesting case of weighted ordinary least-squares is to consider an exponential weight function

$$g(b)=e^{-p|b|} \tag{109}$$

and minimize the error function

$$E=e^{-p|b|}\cdot\frac{1}{N}\sum_{i=1}^{N}\left(a+bx_i-y_i\right)^2 \tag{110}$$

or

$$\begin{aligned}E &= e^{-p|b|}\cdot\left(\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2\right.\\ &\quad\left.+\left(a+b\mu_x-\mu_y\right)^2\right).\end{aligned} \tag{111}$$

for suitable values of p. Eliminate a as usual and solve

$$\frac{d}{db}\left\{e^{-p|b|}\cdot\left(\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2\right)\right\}=0. \tag{112}$$

The result is

$$b=\rho\frac{\sigma_y}{\sigma_x}+\frac{\operatorname{sgn} b}{p}\left(1-\sqrt{1-\left(1-\rho^2\right)\left(\frac{\sigma_y}{\sigma_x}\right)^2p^2}\right) \tag{113}$$

where $\operatorname{sgn} b = \operatorname{sgn}\rho$ as usual. This equation can rewritten as

$$p\left|b-\rho\frac{\sigma_y}{\sigma_x}\right|=1-\sqrt{1-\left(1-\rho^2\right)\left(\frac{\sigma_y}{\sigma_x}\right)^2p^2}. \tag{114}$$

Since the expression inside the radical cannot be negative, this forces a restriction on the admissible values for p, namely:

$$0\le p\le p_0 \tag{115}$$

where

$$p_0=\frac{\sigma_x}{\sigma_y}\frac{1}{\sqrt{1-\rho^2}}. \tag{116}$$

The Hessian matrix

$$H=2e^{-p|b|}\left[\begin{array}{cc}1 & \mu_x\\ \mu_x & \sigma_x^2\left(1-p\left|b-\rho\frac{\sigma_y}{\sigma_x}\right|\right)+\mu_x^2\end{array}\right] \tag{117}$$

is positive definite since

$$\begin{aligned}\det H &= 4\sigma_x^2e^{-2p|b|}\left(1-p\left|b-\rho\frac{\sigma_y}{\sigma_x}\right|\right)\\ &= 4\sigma_x^2e^{-2p|b|}\sqrt{1-\left(1-\rho^2\right)\left(\frac{\sigma_y}{\sigma_x}\right)^2p^2}\end{aligned} \tag{118}$$

which is defined and positive for $0\le p<p_0$.

When $p=0$ the regression degenerates into ordinary least-squares. When $p=p_0$ the coefficient formulas produce the extremal line $y=a_0+b_0x$ where $a_0=\mu_y-b_0\mu_x$ and

$$b_0=\rho\frac{\sigma_y}{\sigma_x}+\operatorname{sgn}\rho\sqrt{1-\rho^2}\frac{\sigma_y}{\sigma_x}. \tag{119}$$

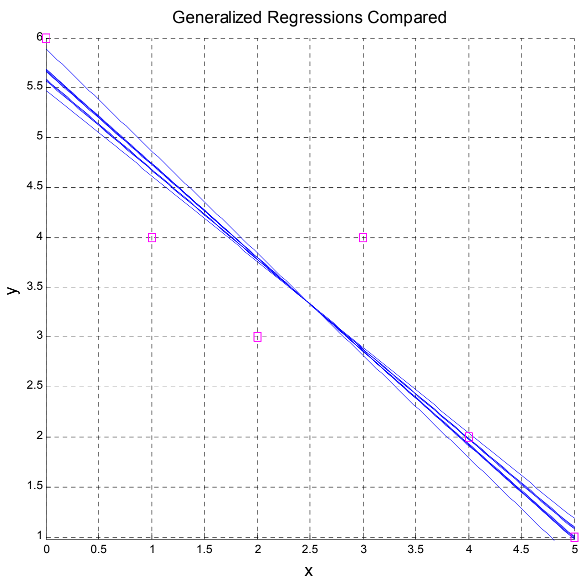
However, $\det H=0$ when $p=p_0$ so that the extremal line's coefficients do not minimize the error function. Nevertheless, it is a useful line to compute since all exponential regression lines are bounded between the OLS $y|x$ line and the extremal line.

4 Numerical Examples

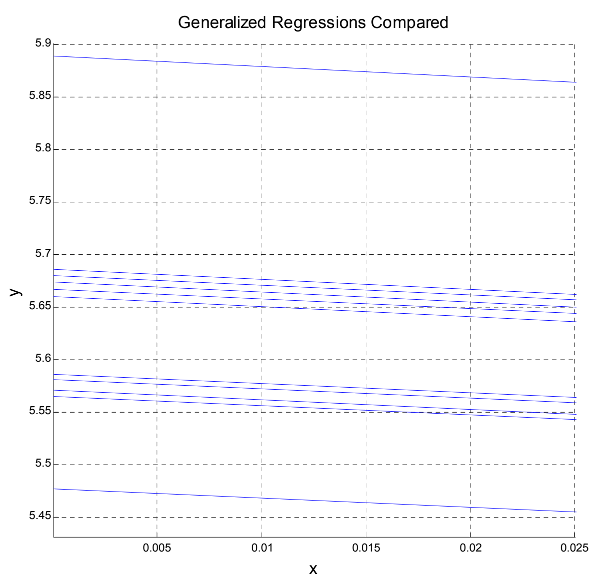
In the examples which follow, all the different regression lines are plotted simultaneously for the same set of data. All generalized least-squares lines $y=a+bx$ pass through the mean point (μ_x,μ_y) since for all the lines $a=\mu_y-b\mu_x$. The outermost regression lines are the ordinary least-squares $x|y$ and $y|x$ lines. In all cases the symmetric and hybrid symmetric lines fan out from the mean point and occupy the space between the ordinary least-squares lines.

ISBN: 978-960-474-319-3

Example 1 Six equispaced data values are considered: $(0,6)$, $(1,4)$, $(2,3)$, $(3,4)$, $(4,2)$, $(5,1)$. The data in this example are taken from Martin's study of orthogonal regression [4] where orthogonal regression was compared with OLS $y|x$ regression. Here all regression lines are compared simultaneously.



The second graph is the same as the first with the region around the y-intercepts magnified.



The regression lines are listed in order of decreasing y-intercept which is how they appear in the graphs.

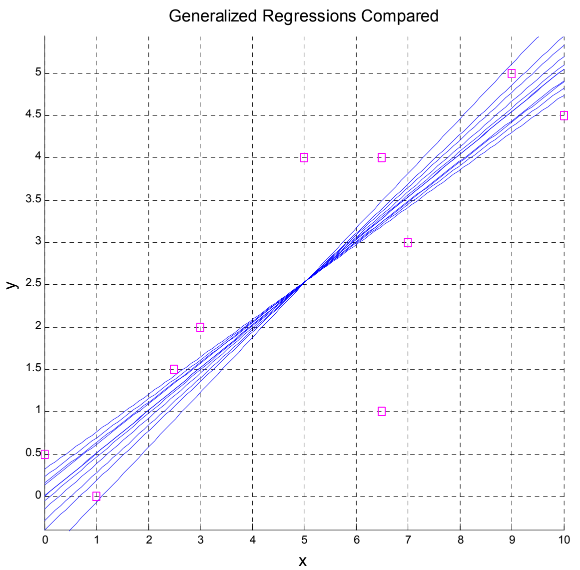
Regression Type	**Equation** $y = a + bx$
OLS $x \mid y$	$y = 5.8889 - 1.02220x$
Pythagorean	$y = 5.6855 - 0.94087x$
Least Perimeter Squared	$y = 5.6797 - 0.93855x$
GMR	$y = 5.6735 - 0.93605x$
Squared Harmonic Mean	$y = 5.6667 - 0.93333x$
Orthogonal	$y = 5.6593 - 0.93038x$
Hybrid Pythagorean	$y = 5.5860 - 0.90107x$
Hybrid Least Perimeter	$y = 5.5811 - 0.89909x$
Hybrid Harmonic Mean	$y = 5.5705 - 0.89486x$
Hybrid Orthogonal	$y = 5.5648 - 0.89260x$
OLS $y \mid x$	$y = 5.4762 - 0.85714x$

From the graphs and the equations it is evident that the symmetric regressions and the hybrid symmetric regressions are banded in separate groups. The symmetric regressions are closer to the OLS $x|y$ line and the hybrid symmetric regressions are closer to the OLS $y|x$ line.

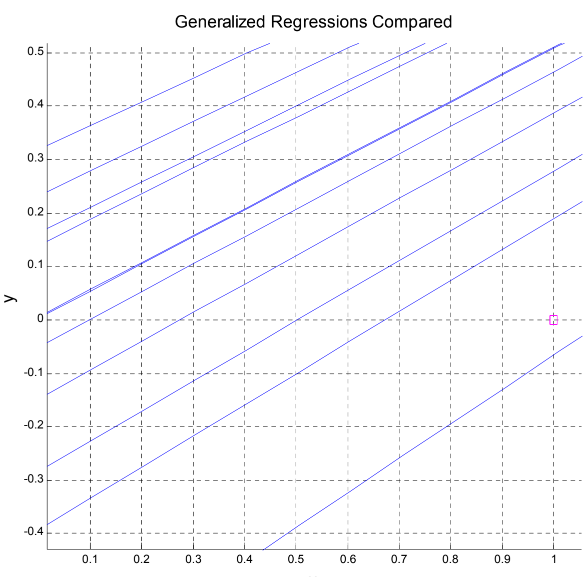
The exponentially-weighted regressions are treated here separately. In this example the admissible values for p are: $0 \leq p \leq p_0$ where $p_0 = 2.6584$. Recall that $p = 0$ reduces to OLS $y|x$ regression and $p = p_0$ is the extremal case. When $p = \frac{1}{2}p_0$ the equation is $y = 5.7282 - 0.95793x$ and when $p = p_0$ the extremal line is $y = 6.4166 - 1.12333x$. All the regression lines are bounded between the OLS $y|x$ line and the extremal line.

Example 2 Ten data values are considered: $(0, 0.5)$, $(1, 0)$, $(2.5, 1.5)$, $(3, 2)$, $(5, 4)$, $(6.5, 4)$, $(6.5, 1)$, $(7, 3)$, $(9, 5)$, $(10, 4.5)$. Like before they are plotted in a scatter plot and all the symmetric and hybrid symmetric regression lines are superimposed.

ISBN: 978-960-474-319-3



The next graph is a magnification of the previous graph around the y-intercepts.



The equations for the lines are again listed in order of decreasing y-intercept, as they appear on the graphs.

Regression Type	**Equation** $y = a + bx$
OLS $y \mid x$	$y = 0.32018 + 0.44155x$
Hybrid Orthogonal	$y = 0.23357 + 0.45870x$
Hybrid Harmonic Mean	$y = 0.16383 + 0.47251x$
Orthogonal	$y = 0.14064 + 0.47710x$
Hybrid Least Perimeter	$y = 0.0065259 + 0.50360x$
Squared Harmonic Mean	$y = 0.0041322 + 0.50413x$
Hybrid Pythagorean	$y = -0.050365 + 0.51492x$
GMR	$y = -0.14682 + 0.53402x$
Least Perimeter Squared	$y = -0.28237 + 0.56086x$
Pythagorean	$y = -0.39241 + 0.58266x$
OLS $x \mid y$	$y = -0.71164 + 0.64587x$

In this example, the graphs and the table also suggest a tendency for symmetric and hybrid symmetric regression lines to remain separate, however in this case there is some overlap.

Here the admissible parameter values for exponentially-weighted regressions are: $0 \le p \le p_0$, where $p_0 = 3.3293$. Again $p = 0$ corresponds to OLS $y|x$. When $p = \frac{1}{2}p_0$ the regression line is $y = -0.086254 + 0.52203x$. When $p = p_0$ the regression line is extremal and its equation is $y = -1.1966 + 0.74191x$. All the lines are bounded between the OLS $y|x$ line and the extremal line.

In these examples the correlation coefficient ρ is close in absolute value to 1 and so $\frac{1}{\rho}$ is also close in absolute value to 1. As a result the OLS $y|x$ and $x|y$ lines are visually near each other and both lines are seen to pass through the data. However, whenever the correlation coefficient is small, the OLS $x|y$ line will diverge noticeably from the other lines and from the data as well. This is not shown here.

5 Summary

The derivation of least-squares regressions involves constructing the summation expression for the mean squared error between the data and the line, denoted here by E. In the standard derivation, E_a and E_b are set equal to zero, and the equations are solved for minimizing solution (a, b). To check that the solution is actually a minimum, the Hessian determinant must be computed and found to be positive.

The approach of this paper has several notable features which made the derivations of generalized

ISBN: 978-960-474-319-3

least-squares regressions as efficient and uncomplicated as possible. First, this paper utilizes a high-level formula for the ordinary least-squares error E due to Ehrenberg which is already expressed in terms of μ_x, μ_y, σ_x, σ_y and ρ. This formula is not widely-known but it deserves to be. The formula allows for an elementary derivation of ordinary least-squares and it allows all the generalized least-squares methods described here to be derived without complicated summation manipulations. In all cases, the computations for E_a and E_b become calculus-level problems. The calculation of the Hessian matrix and determinant becomes tractable now in all cases as well.

A second notable feature in all the derivations is the use of a relation expressing the deviation between a data point and the line with respect to x in terms of the deviation of the data point with respect to y. It is used here in every derivation to extract a weight function $g(b)$ from the error expression E. In this way, the generalized least-squares error is always a product of the weight function and Ehrenberg's formula for ordinary least-squares error.

With the pattern of derivation now simplified and streamlined in the known cases, we were able to explore a variety of new generalized least-squares methods in the same efficient manner. Whenever possible, explicit formulas for the new regression coefficients a and b were determined. In those cases where the slope was a solution to a cubic or quartic equation, Descartes' Rule of Signs and sometimes the polynomial discriminant were invoked to prove the existence of a unique solution having the same sign as ρ.

Numerical examples show the symmetric and hybrid symmetric regression lines fanning out from the mean point and occupying the space between the ordinary least-squares lines. The examples also reveal a tendency for the symmetric regression lines to remain separate from the hybrid regression lines. In future work, the accuracy of these regression lines will be analyzed.

The work of this paper is generalized into a theory for deriving and classifying generalized least-squares regressions in the paper to follow.

References

[1] A. S. C. Ehrenberg, Deriving the Least-Squares Regression Equation, *The American Statistician*, Vol. 37, No. 3 (Aug. 1983), p.232.

[2] V. Ferlini, Mathematics without (many) words, *College Math Journal*, No. 33, (2002), p. 170.

[3] N. Greene, Generalized Least-Squares Regressions II: Theory and Classification, *Proceedings of the 1st International Conference on Computational Science and Engineering (CSE '13)*, Valencia, Spain, August 6-8, 2013.

[4] 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), pp.1-5.

[5] S. B. Martin, *Less than the Least: An Alternative Method of Least-Squares Linear Regression*, Undergraduate Honors Thesis, Department of Mathematics, McMurry University, Abilene, Texas, 1998.

[6] R. B. Nelson, Proof Without Words: A Reciprocal Pythagorean Theorem, *Mathematics Magazine*, Vol. 82, No. 5, Dec. 2009, p.370.

[7] P. A. Samuelson, A Note on Alternative Regressions, *Econometrica*, Vol. 10, No. 1 (Jan. 1942) pp. 80-83.

[8] R. Taagepera, *Making Social Sciences More Scientific: The Need for Predictive Models*, Oxford University Press, New York, 2008.

[9] E.W. Weisstein, Discriminant (Polynomial) in: *CRC Concise Encyclopedia of Mathematics*, 2nd. Edition, Chapman & Hall/CRC Boca Raton, 2003, p. 784.

[10] E. B. Woolley, The Method of Minimized Areas as a Basis for Correlation Analysis, *Econometrica*, Vol. 9, No. 1, (Jan 1941), pp.38-62.

ISBN: 978-960-474-319-3