2013

Generalized Least-Squares Regressions II: Theory and Classification

Nataniel Greene
CUNY Kingsborough Community College

Generalized Least-Squares Regressions II: Theory and Classification

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: In the first paper of this series, a variety of known and new symmetric and weighted least-squares regression methods were presented with efficient derivations. This paper continues and generalizes the previous work with a theory for deriving, analyzing, and classifying all symmetric and weighted least-squares regression methods.

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

1 Overview

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. Two alternative symmetric regression methods which overcome this problem are orthogonal regression and geometric mean regression. In the first paper of this series [2], a variety of alternative symmetric and weighted regression methods were presented and analyzed. The derivations were efficient in their use of Ehrenberg's formula for the ordinary least-squares error [1], avoiding the cumbersome algebraic manipulations with summations which would otherwise have been necessary. The derivations also extracted a unique weight function $g(b)$ from the error expression in every case. Clearly there are infinitely many cases of symmetric and weighted regressions, and a similar efficient derivation will apply in every case.

With the pattern of derivation now clearly established, this paper generalizes the procedures in a theory for computing and classifying any generalized least-squares regression. In the theory, the general symmetric least-squares problem and the general weighted least-squares problem are formally defined. Since generalized regressions are characterized by their weight function $g(b)$, this paper derives formulas for the regression coefficients a and b in terms of $g(b)$ as well as general formulas for the Hessian matrix and determinant. In the process, formulas for general classes of weight functions emerge, and all the regression cases derived previously are categorized as belonging to various weight function classes.

2 Theory and Classification

The theory for classifying least-squares regressions is detailed now.

2.1 General Framework and Coefficient Formulas

Definition 1 *Define a function $\psi(x,y)$ that is:*
(i) Non-negative:

$$\psi(x,y) \geq 0$$

(ii) Symmetric in x and y:

$$\psi(x,y) = \psi(y,x)$$

(iii) Even in x and y:

$$\psi(-x,-y) = \psi(x,y)$$

(iv) Non-decreasing in x and y. For ψ differentiable this means

$$\psi_x \geq 0 \text{ and } \psi_y \geq 0.$$

(v) Homogeneous with degree 2 in x and y:

$$\psi(\lambda x, \lambda y) = \lambda^2 \psi(x,y).$$

Definition 2 *(The General Symmetric Least-Squares Problem) Values of a and b are sought which minimize an error function defined by*

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

Definition 3 *(The General Weighted Ordinary Least-Squares Problem) Values of a and b are sought which minimize an error function defined by*

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

or

$$\begin{aligned} E &= g(b)\left(\sigma_y^2\left(1-\rho^2\right) + (b\sigma_x - \rho\sigma_y)^2 \right.\\ &\quad \left. + \left(a + b\mu_x - \mu_y\right)^2\right) \quad (3)\end{aligned}$$

where $g(b)$ is a positive even function that is non-decreasing for $b < 0$ and non-increasing for $b > 0$.

The next theorem is fundamental and it is already anticipated from the previous work. It states that *every generalized symmetric least-squares problem is equivalent to a weighted ordinary least-squares problem* with weight function given by $g(b)$.

Theorem 4 *The general symmetric least-squares error function can be written equivalently as*

$$E = g(b)\cdot\frac{1}{N}\sum_{i=1}^{N}(a + bx_i - y_i)^2 \quad (4)$$

or

$$\begin{aligned} E &= g(b)\left(\sigma_y^2\left(1-\rho^2\right) + (b\sigma_x - \rho\sigma_y)^2 \right.\\ &\quad \left. + \left(a + b\mu_x - \mu_y\right)^2\right) \quad (5)\end{aligned}$$

where

$$g(b) = \psi\left(1, \frac{1}{b}\right). \quad (6)$$

Proof. Substitute $\frac{a}{b} + x_i - \frac{1}{b}y_i$ with $\frac{1}{b}(a + bx_i - y_i)$ and then use the homogeneity property:

$$\begin{aligned} E &= \frac{1}{N}\sum_{i=1}^{N}\psi\left(a + bx_i - y_i, \frac{a}{b} + x_i - \frac{1}{b}y_i\right)\\ &\quad \frac{1}{N}\sum_{i=1}^{N}\psi\left(a + bx_i - y_i, \frac{1}{b}(a + bx_i - y_i)\right)\\ &= \frac{1}{N}\sum_{i=1}^{N}(a + bx_i - y_i)^2\,\psi\left(1, \frac{1}{b}\right).\end{aligned}$$

Define

$$g(b) = \psi\left(1, \frac{1}{b}\right)$$

and write

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

Substitute using Ehrenberg's formula and obtain

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

■

From the characterization of the generating function it follows that the weight function must be a positive even function of b that is non-decreasing for $b < 0$ and non-increasing for $b > 0$. The following table summarizes the non-hybrid symmetric regressions detailed earlier.

Case	**Generating Function** $\psi(x,y)$	**OLS Weight Function** $g(b)$
OLS $y \mid x$	x^2	1
OLS $x \mid y$	y^2	$\frac{1}{b^2}$
Orthogonal	$\frac{x^2y^2}{x^2+y^2}$	$\frac{1}{1+b^2}$
GMR	$\lvert xy\rvert$	$\frac{1}{\lvert b\rvert}$
Pythagorean	x^2+y^2	$1+\frac{1}{b^2}$
Least Perimeter Squared	$(\lvert x\rvert+\lvert y\rvert)^2$	$\left(1+\frac{1}{\lvert b\rvert}\right)^2$
Squared Harmonic Mean	$\left(\frac{\lvert xy\rvert}{\lvert x\rvert+\lvert y\rvert}\right)^2$	$\frac{1}{(1+\lvert b\rvert)^2}$

Theorem 5 *Every generating function $\psi(x,y)$ can be recovered from its corresponding weight function $g(b)$ using*

$$\psi(x,y) = x^2 g\left(\frac{x}{y}\right) = y^2 g\left(\frac{y}{x}\right). \quad (7)$$

Proof. By definition and homogeneity

$$\begin{aligned} x^2 g\left(\frac{x}{y}\right) &= x^2\psi\left(1, \frac{1}{x/y}\right)\\ &= \psi\left(x, \frac{x}{x/y}\right)\\ &= \psi(x,y)\end{aligned}$$

ISBN: 978-960-474-319-3

and the second formula follows by symmetry. ■

The procedure for finding a and b is to again set partial derivatives of the error function equal to zero and then solve the resulting equations for a and b. This is done in the next theorem for the general case. The result is a general equation for the slope b in terms of ρ, σ_x, σ_y, $g(b)$ and $g'(b)$ which can be used to verify the specific formulas for b in the first paper with greater ease. The formula for b is called here the First Discrepancy Formula because the left hand side is the discrepancy between the generalized least-squares coefficient b and the ordinary least-squares coefficient given by $\rho\frac{\sigma_y}{\sigma_x}$.

Theorem 6 *The y-intercept a and the slope b of the generalized least-squares regression line satisfy:*

$$a = \mu_y - b\mu_x \tag{8}$$

and

$$\frac{d}{db}\left\{g(b)\left(\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2\right)\right\}=0. \tag{9}$$

Proof. Begin with the error function

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

Take the first partial derivative with respect to a and set it equal to zero.

$$E_a = 2g(b)\left(a+b\mu_x-\mu_y\right)=0$$

Solve for a and obtain $a = \mu_y - b\mu_x$.

Next, take the first partial derivative of the error function with respect to b and set it equal to zero.

$$\begin{aligned} 0 &= E_b \\ &= g'(b)\left(\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2\right. \\ &\quad \left.+\left(a+b\mu_x-\mu_y\right)^2\right) \\ &\quad +2g(b)\left(\sigma_x\left(b\sigma_x-\rho\sigma_y\right)+\mu_x\left(a+b\mu_x-\mu_y\right)\right) \end{aligned}$$

Substitute $a = \mu_y - b\mu_x$

$$\begin{aligned} 0 &= g'(b)\left(\sigma_y^2\left(1-\rho^2\right)+\left(b\sigma_x-\rho\sigma_y\right)^2\right) \\ &\quad +2g(b)\,\sigma_x\left(b\sigma_x-\rho\sigma_y\right). \end{aligned}$$

Observe that the same result is obtained by first substituting $a = \mu_y - b\mu_x$ and eliminating a from the error function. Then one can solve the equation

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

for b. This is the simpler computation to perform. ■

Theorem 7 *(First Discrepancy Formula) The discrepancy between the generalized least-squares coefficient b and the ordinary least-squares coefficient $\rho\frac{\sigma_y}{\sigma_x}$ is given implicitly by*

$$\begin{aligned} b-\rho\frac{\sigma_y}{\sigma_x} &= -\frac{1}{2}\frac{g'(b)}{g(b)}\left(\left(b-\rho\frac{\sigma_y}{\sigma_x}\right)^2\right. \\ &\quad \left.+\left(\frac{\sigma_y}{\sigma_x}\right)^2\left(1-\rho^2\right)\right) \end{aligned} \tag{10}$$

Proof. Perform the differentiation indicated by the previous formula. ■

Corollary 8 *The general regression coefficient b always has the same sign as ρ. Denote the ordinary least-squares coefficient $\rho\frac{\sigma_y}{\sigma_x}$ by b_{OLS}. When ρ is positive $b > b_{OLS}$ and when ρ is negative $b < b_{OLS}$.*

Proof. The above equation for b calculates the discrepancy between ordinary least squares slope $\rho\frac{\sigma_y}{\sigma_x}$ and the least squares slope based on the function g. Observe that the expression

$$\left(b-\rho\frac{\sigma_y}{\sigma_x}\right)^2+\left(\frac{\sigma_y}{\sigma_x}\right)^2\left(1-\rho^2\right)$$

is always positive since $|\rho| \le 1$. The function $g(b)$ is always positive and $g'(b)$ is always negative for b positive and $g'(b)$ is always positive for b negative. Conclude that when $b > 0$, the right hand side is positive and therefore $b > \rho\frac{\sigma_y}{\sigma_x}$. Similarly, when b is negative the right hand side is negative and $b < \rho\frac{\sigma_y}{\sigma_x}$. This implies that $\operatorname{sgn} b = \operatorname{sgn} \rho$. ■

The First Discrepancy Formula is useful for deriving the formulas for b in the cases that were already worked out, but it is also problematic in that it is an implicit formula in $b-\rho\frac{\sigma_y}{\sigma_x}$. In the next theorem an explicit formula for $b-\rho\frac{\sigma_y}{\sigma_x}$ is given.

Theorem 9 *(Second Discrepancy Formula) An explicit formula for the discrepancy is given by*

$$\begin{aligned} b-\rho\frac{\sigma_y}{\sigma_x} &= -\frac{g(b)}{g'(b)} \\ &\quad -\operatorname{sgn}\rho\sqrt{\left(\frac{g(b)}{g'(b)}\right)^2-\left(\frac{\sigma_y}{\sigma_x}\right)^2\left(1-\rho^2\right)} \end{aligned} \tag{11}$$

Proof. Use the quadratic formula to solve for $b-\rho\frac{\sigma_y}{\sigma_x}$. Choose the sign in front of the radical to be $-\operatorname{sgn}\rho$. ■

Lemma 10 *The following inequality is true for all b for which the expression is defined*

$$\left|\frac{g'(b)}{g(b)}\right| \leq \frac{\sigma_x}{\sigma_y}\frac{1}{\sqrt{1-\rho^2}} \tag{12}$$

Proof. The quantity under the radical in the second discrepancy formula must necessarily be nonnegative. Therefore

$$\begin{aligned}\left(\frac{g(b)}{g'(b)}\right)^2 &\geq \left(\frac{\sigma_y}{\sigma_x}\right)^2\left(1-\rho^2\right)\\ \left(\frac{g'(b)}{g(b)}\right)^2 &\leq \left(\frac{\sigma_x}{\sigma_y}\right)^2\frac{1}{1-\rho^2}\\ \left|\frac{g'(b)}{g(b)}\right| &\leq \frac{\sigma_x}{\sigma_y}\frac{1}{\sqrt{1-\rho^2}}.\end{aligned}$$

■

It is shown now that $g(b)$ can grow or decay at most exponentially over an interval.

Theorem 11 *Let $g(b)$ be defined over an interval $[b_0, b]$. Then $g(b)$ can grow or decay at most exponentially over this interval. More specifically*

$$ke^{-b\frac{\sigma_x}{\sigma_y}\frac{1}{\sqrt{1-\rho^2}}} \leq g(b) \leq Ke^{b\frac{\sigma_x}{\sigma_y}\frac{1}{\sqrt{1-\rho^2}}} \tag{13}$$

where

$$k = g(b_0)\,e^{b_0\frac{\sigma_x}{\sigma_y}\frac{1}{\sqrt{1-\rho^2}}}$$

and

$$K = g(b_0)\,e^{-b_0\frac{\sigma_x}{\sigma_y}\frac{1}{\sqrt{1-\rho^2}}}.$$

Proof. Rewrite the previous inequality as

$$-\frac{\sigma_x}{\sigma_y}\frac{1}{\sqrt{1-\rho^2}} \leq \frac{g'(b)}{g(b)} \leq \frac{\sigma_x}{\sigma_y}\frac{1}{\sqrt{1-\rho^2}}$$

or

$$-\frac{\sigma_x}{\sigma_y}\frac{1}{\sqrt{1-\rho^2}} \leq \frac{d}{da}\ln g(b) \leq \frac{\sigma_x}{\sigma_y}\frac{1}{\sqrt{1-\rho^2}}.$$

Integrate all three sides of the inequality over $[b_0, b]$ and obtain

$$-\frac{\sigma_x}{\sigma_y}\frac{b-b_0}{\sqrt{1-\rho^2}} \leq \ln g(b) - \ln g(b_0) \leq \frac{\sigma_x}{\sigma_y}\frac{b-b_0}{\sqrt{1-\rho^2}}.$$

Exponentiate all three sides, solve for $g(b)$ and obtain the inequality. ■

2.2 The Weight Function and Relative Error

It was seen that the error function for generalized least-squares regression is the product of the weight function $g(b)$ and the ordinary least-squares error function. It follows that the weight function $g(b)$ is the ratio of generalized least-squares error to ordinary least-squares error:

$$g(b) = \frac{E_g}{E_{\text{OLS}}}. \tag{14}$$

For several regressions the weight function $g(b)$ is strictly less than 1. In these cases

$$E_g < E_{\text{OLS}} \tag{15}$$

which may be a desirable feature for a regression to have.

If ordinary least-squares regression is viewed as the standard, then the relative error between the two regression errors is given by

$$\frac{|E_{\text{OLS}} - E_g|}{E_{\text{OLS}}} = |1 - g(b)|. \tag{16}$$

If the generalized least-squares regression is viewed as the standard, then the relative error is given by

$$\frac{|E_g - E_{\text{OLS}}|}{E_g} = \left|1 - \frac{1}{g(b)}\right|. \tag{17}$$

More generally, for any two regressions with weight functions $g_1(b)$ and $g_2(b)$ the error relative to the $g_2(b)$ regression is given by

$$\frac{|E_{g_2} - E_{g_1}|}{E_{g_2}} = \left|1 - \frac{g_1(b)}{g_2(b)}\right|. \tag{18}$$

Multiplying by 100% yields the equivalent percent errors. The methods can be ranked according to the percent error relative to ordinary least-squares regression, relative to geometric mean regression, or relative to any other generalized regression. In this way the fundamental role of the weight function in analyzing generalized regressions is further underscored.

2.3 The Indicative Function and the Hessian Matrix

In order for the values of a and b to minimize the error function, the Hessian matrix of second order partial derivatives evaluated at a and b must also be positive definite. The general Hessian matrix is calculated in the next theorem.

For a given weight function $g(b)$ a particular combination of g and its first and second derivatives always occurs in the calculation of H and $\det H$. It is denoted here by $G(b)$. It plays a fundamental role in indicating whether the Hessian matrix will be positive definite. It also plays a fundamental role in determining the common form which all weight functions possess.

Definition 12 *Define the indicative function*

$$G(b) = \frac{2g'(b)}{g(b)} - \frac{g''(b)}{g'(b)}. \tag{19}$$

and call the equation the indicative equation.

Theorem 13 *The general Hessian matrix can be written compactly as*

$$H = 2g(b) \begin{bmatrix} 1 & \mu_x \\ \mu_x & \frac{\det H}{4g(b)^2} + \mu_x^2 \end{bmatrix} \tag{20}$$

where

$$\det H = 4g(b)^2 \sigma_x^2 \left\{ 1 + G(b) \left(b - \rho \frac{\sigma_y}{\sigma_x} \right) \right\}. \tag{21}$$

Proof. Begin with

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

and take second order partial derivatives:

$$\begin{aligned} E_{aa} &= \frac{\partial}{\partial a} \left(2g(b)\left(a + b\mu_x - \mu_y\right)\right) \\ &= 2g(b) \end{aligned}$$

and

$$\begin{aligned} E_{ab} &= \frac{\partial}{\partial b} \left(2g(b)\left(a + b\mu_x - \mu_y\right)\right) \\ &= 2g(b)\mu_x + 2g'(b)\left(a + b\mu_x - \mu_y\right). \end{aligned}$$

Replace a with $\mu_y - b\mu_x$ and obtain

$$E_{ab} = E_{ba} = 2g(b)\mu_x.$$

The second derivative with respect to b is given by

$$\begin{aligned} E_{bb} &= \frac{\partial}{\partial b} \{2g(b)\left(\sigma_x\left(b\sigma_x - \rho\sigma_y\right)\right. \\ &\quad \left. + \mu_x\left(a + b\mu_x - \mu_y\right)\right) \\ &\quad + g'(b)\left(\sigma_y^2\left(1-\rho^2\right) + \left(b\sigma_x - \rho\sigma_y\right)^2\right. \\ &\quad \left.\left. + \left(a + b\mu_x - \mu_y\right)^2\right)\right\} \\ &= 2g'(b)\left(\sigma_x\left(b\sigma_x - \rho\sigma_y\right) + \mu_x\left(a + b\mu_x - \mu_y\right)\right) \\ &\quad + g''(b)\left(\sigma_y^2\left(1-\rho^2\right) + \left(b\sigma_x - \rho\sigma_y\right)^2\right. \\ &\quad \left. + \left(a + b\mu_x - \mu_y\right)^2\right) \\ &\quad + 2g(b)\left(\sigma_x^2 + \mu_x^2\right) + 2g'(b)\sigma_x\left(b\sigma_x - \rho\sigma_y\right) \\ &\quad + \mu_x\left(a + b\mu_x - \mu_y\right). \end{aligned}$$

Replace a with $\mu_y - b\mu_x$ and obtain

$$\begin{aligned} E_{bb} &= 2g(b)\left(\sigma_x^2 + \mu_x^2\right) + 4g'(b)\left(b\sigma_x^2 - 2\rho\sigma_x\sigma_y\right) \\ &\quad + g''(b)\left(\left(b\sigma_x - \rho\sigma_y\right)^2 + \sigma_y^2\left(1-\rho^2\right)\right)\Big\} \\ &= 2g(b)\left\{\left(\sigma_x^2 + \mu_x^2\right) + \sigma_x^2 \frac{2g'(b)}{g(b)}\left(b - \rho\frac{\sigma_y}{\sigma_x}\right)\right. \\ &\quad + \sigma_x^2 \frac{g''(b)}{g(b)}\left(\left(b - \rho\frac{\sigma_y}{\sigma_x}\right)^2\right. \\ &\quad \left.\left. + \left(\frac{\sigma_y}{\sigma_x}\right)^2\left(1-\rho^2\right)\right)\right\}. \end{aligned}$$

Now use the First Discrepancy Formula to replace $\left(b - \rho\frac{\sigma_y}{\sigma_x}\right)^2 + \left(\frac{\sigma_y}{\sigma_x}\right)^2\left(1-\rho^2\right)$ with $-2\frac{g(b)}{g'(b)}\left(b - \rho\frac{\sigma_y}{\sigma_x}\right)$ and obtain

$$\begin{aligned} E_{bb} &= 2g(b)\left\{\left(\sigma_x^2 + \mu_x^2\right)\right. \\ &\quad \left. + \sigma_x^2\left(\frac{2g'(b)}{g(b)} - \frac{g''(b)}{g'(b)}\right)\left(b - \rho\frac{\sigma_y}{\sigma_x}\right)\right\}. \end{aligned}$$

Finally

$$\begin{aligned} \det H &= E_{aa}E_{bb} - E_{ab}^2 \\ &= 4\left(g(b)\right)^2\left\{\left(\sigma_x^2 + \mu_x^2\right)\right. \\ &\quad \left. + \sigma_x^2\left(\frac{2g'(b)}{g(b)} - \frac{g''(b)}{g'(b)}\right)\left(b - \rho\frac{\sigma_y}{\sigma_x}\right)\right\} \\ &\quad - 4\left(g(b)\right)^2\mu_x^2 \\ &= 4\left(g(b)\right)^2\sigma_x^2\left\{1 + \left(\frac{2g'(b)}{g(b)} - \frac{g''(b)}{g'(b)}\right)\right. \\ &\quad \left. \times\left(b - \rho\frac{\sigma_y}{\sigma_x}\right)\right\}. \end{aligned}$$

■

ISBN: 978-960-474-319-3

The next theorem is a simple corollary of the previous theorem. However, it is called a theorem in order that the reader not miss its significance. It gives the simplest way to check for positive-definiteness: use the indicative function.

Theorem 14 *Suppose $g(b)$ and $g'(b)$ are not zero. Then the Hessian matrix is positive definite and the Hessian determinant is positive if and only if*

$$G(b)\left(b-\rho\frac{\sigma_y}{\sigma_x}\right) > -1. \tag{22}$$

Any function $G(b)$ satisfying this condition is admissible as an indicative function. For example, any $G(b)$ such that $\operatorname{sgn} G(b) = \operatorname{sgn} b$ is admissible as an indicative function, since $\operatorname{sgn} b = \operatorname{sgn} \rho$ ensures that $G(b)\left(b-\rho\frac{\sigma_y}{\sigma_x}\right) > 0$. More generally, any $G(b)$ of the form $G(b) = p(b) - \frac{1}{b-\rho\frac{\sigma_y}{\sigma_x}}$ where $p(b)$ is a positive function is admissible.

To aid the reader in the process of verifying the Hessian determinant formulas presented earlier, a table of indicative functions is presented.

Case	**Weight Function** $g(b)$	**Indicative Function** $G(b)$
OLS $y\,\vert\,x$	1	NA
OLS $x\,\vert\,y$	$\frac{1}{b^2}$	$-\frac{1}{b}$
Orthogonal	$\frac{1}{1+b^2}$	$-\frac{1}{b}$
GMR	$\frac{1}{\vert b\vert}$	0
Pythagorean	$1+\frac{1}{b^2}$	$-\frac{1}{b}+\frac{4b}{b^2+1}$
Least Perimeter Squared	$\left(1+\frac{1}{\vert b\vert}\right)^2$	$-\frac{1}{b}+\frac{3\operatorname{sgn} b}{\vert b\vert+1}$
Squared Harmonic Mean	$\frac{1}{(1+\vert b\vert)^2}$	$-\frac{\operatorname{sgn} b}{\vert b\vert+1}$
Hybrid Least Perimeter	$1+\frac{1}{\vert b\vert}$	$\frac{2\operatorname{sgn} b}{\vert b\vert+1}$
Hybrid Harmonic Mean	$\frac{1}{1+\vert b\vert}$	0
Hybrid Pythagorean	$\sqrt{1+\frac{1}{b^2}}$	$\frac{3b}{b^2+1}$
Hybrid Orthogonal	$\frac{1}{\sqrt{1+b^2}}$	$-\frac{1}{b}+\frac{b}{b^2+1}$
Exponential	$\exp(-p\vert b\vert)$	$-p\operatorname{sgn} b$

There is clearly redundancy in this table. It is apparent, for example, that different weight functions can have the same indicative function. Furthermore, several indicative functions differ from each other only by a multiplicative constant. To consolidate and further generalize the above table, the inverse problem of determining all weight functions corresponding to a given indicative function must be solved. The solution to the inverse problem reveals the common form that all the weight functions share.

Theorem 15 *Let $G(b)$ be an indicative function. Then a general solution $g(b)$ to the indicative equation is given by:*

$$g(b) = \frac{1}{c + k\int \exp\left(-\int G(b)\,db\right)db} \tag{23}$$

where c and k are arbitrary constants.

Proof. The indicative equation

$$G(b) = \frac{2g'(b)}{g(b)} - \frac{g''(b)}{g'(b)}$$

is solved for $g(b)$. Write

$$\begin{aligned} G(b) &= \frac{2g'(b)}{g(b)} - \frac{g''(b)}{g'(b)} \\ &= 2\frac{d}{db}\ln|g(b)| - \frac{d}{db}\ln|g'(b)| \\ &= \frac{d}{db}\ln\left(\frac{(g(b))^2}{|g'(b)|}\right). \end{aligned}$$

Therefore

$$\ln\left(\frac{(g(b))^2}{|g'(b)|}\right) = \int G(b)\,db + C$$

and

$$\frac{(g(b))^2}{g'(b)} = K\exp\left(\int G(b)\,db\right)$$

where $K = \pm e^C$. The resulting differential equation is now a Bernoulli equation:

$$\frac{g'(b)}{(g(b))^2} = K^{-1}\exp\left(-\int G(b)\,db\right).$$

To solve it, let $u = \frac{1}{g(b)}$ so that $u' = -\frac{g'(b)}{(g(b))^2}$. Substituting for u yields

$$u'(b) = -K^{-1}\exp\left(-\int G(b)\,db\right)$$

and

$$u(b) = k\int \exp\left(-\int G(b)\,db\right)db + c$$

ISBN: 978-960-474-319-3

where $k = -K^{-1}$. Substituting back for $g(b)$ yields the result. ■

The next theorem shows that linear combinations of indicative functions produce valid weight functions.

Theorem 16 *(Linear Combinations of Indicative Functions). Let $G_0(b)$, $G_1(b)$ and $G_2(b)$ be indicative functions with corresponding weight functions:*

$$g_{\{G_0\}}(b) = \frac{1}{c_0 + k_0 \int \gamma_0(b)\, db},$$

$$g_{\{G_1\}}(b) = \frac{1}{c_1 + k_1 \int \gamma_1(b)\, db}$$

and

$$g_{\{G_2\}}(b) = \frac{1}{c_2 + k_2 \int \gamma_2(b)\, db}.$$

Then:

1. The weight functions corresponding to multiplication of $G_0(b)$ by a constant p are given by

$$g_{\{pG_0\}}(b) = \frac{1}{c + k \int (\gamma_0(b))^p\, db}. \tag{24}$$

2. The weight functions corresponding to a sum $G_1(b) + G_2(b)$ are given by

$$g_{\{G_1+G_2\}}(b) = \frac{1}{c + k \int \gamma_1(b)\, \gamma_2(b)\, db}. \tag{25}$$

3. The weight functions corresponding to a difference $G_1(b) - G_2(b)$ are given by

$$g_{\{G_1-G_2\}}(b) = \frac{1}{c + k \int \frac{\gamma_1(b)}{\gamma_2(b)} db}. \tag{26}$$

4. The weight functions corresponding to a general linear combination $pG_1(b) + qG_2(b)$ are given by

$$g_{\{pG_1+qG_2\}}(b) = \frac{1}{c + k \int (\gamma_1(b))^p\, (\gamma_2(b))^q\, db}. \tag{27}$$

With the above theorems, any admissible function $G(b)$ can now be used to construct a weighted ordinary least-squares regression problem. Linear combinations of previously known indicative functions can also be formed and the resulting weight functions are more easily constructed. As always, the value for b which minimizes the error is determined by solving the First Discrepancy Formula for b and setting $a = \mu_y - b\mu_x$.

The table presented next is a consolidation and a generalization of the previous chart. Every regression is categorized using a common indicative function. To every general indicative function there is a corresponding class of weight functions. Many of the regressions derived previously and their weight functions are now seen to be instances of the same general weight function.

Indicative Function $G(b)$	**General Weight Function** $g(b)$	**Specific Cases**
1. 0	$\frac{1}{c+k\lvert b\rvert}$	**GMR:** $c=0, k=1,\ g(b)=1/\lvert b\rvert$ **Hybrid Harmonic Mean:** $c=1, k=1,$ $g(b)=1/(1+\lvert b\rvert)$
2. $\frac{p}{b}$	$\frac{1}{c+k\lvert b\rvert^{1-p}}$	**OLS** $x\vert y$**:** $c=0, k=1, p=-1,$ $g(b)=1/b^2$ **Orthogonal:** $c=1, k=1, p=-1,$ $g(b)=1/(1+b^2)$ **Previous cases:** $p=0$.
3. $-p \operatorname{sgn} b$	$\frac{1}{c+k\exp(p\lvert b\rvert)}$	**Exponential:** $c=0, k=1,\ 0<p<p_0$ $g(b)=\exp(-p\lvert b\rvert)$
4. $\frac{pb}{b^2+1}$	$\frac{1}{c+k\int(b^2+1)^{-p/2}\,db}$	**Hybrid Pythagorean:** $c=0, k=1, p=3,$ $g(b)= 1+1/b^2$
5. $-\frac{1}{b}+\frac{pb}{b^2+1}$	$\frac{1}{c+k(b^2+1)^{1-p/2}}$	**Pythagorean:** $c=1, k=-1, p=4,$ $g(b)=1+1/b^2$ **Hybrid Orthogonal:** $c=0, k=1, p=1,$ $g(b)=1/\sqrt{1+b^2}$
6. $\frac{p \operatorname{sgn} b}{\lvert b\rvert+1}$	$\frac{1}{c+k(\lvert b\rvert+1)^{1-p}}$	**Squared Harmonic Mean:** $c=0, k=1, p=-1,$ $g(b)=1/(1+\lvert b\rvert)^2$ **Hybrid Least Perimeter:** $c=1, k=-1, p=2,$ $g(b)=1+1/\lvert b\rvert$
7. $-\frac{1}{b}+\frac{p \operatorname{sgn} b}{\lvert b\rvert+1}$	$\frac{1}{c+k(\lvert b\rvert+1)^{1-p}((p-1)\lvert b\rvert+1)}$	**Least Perimeter Squared:** $c=1, k=-1, p=3,$ $g(b)=(1+1/\lvert b\rvert)^2$

The table reveals the hidden relationships and the underlying unity behind the disparate regressions presented previously. It also opens up many more specific cases of weighted ordinary least-squares regression for future exploration. Further generalizations and additions to this table are possible as well. The detailed construction of generalized least-squares problems based on other choices for $G(b)$ is a subject for future work.

3 Summary

The derivation of least-squares regressions involves constructing the summation expression for the mean squared error between the data and the line, denoted

ISBN: 978-960-474-319-3

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.

In the first paper of this series, efficient derivations for a variety of generalized least-squares regressions were presented on a case-by-case basis. The novelty of the derivations lied in their use of Ehrenberg's formula for the ordinary least-squares error, avoiding cumbersome algebraic manipulations with summation symbols. The derivations also related the x and y deviations of the data to the slope of the line in order to extract a weight function $g(b)$ from the error expression.

With the pattern of derivation now clearly established, this paper generalizes the procedure into a theory for computing and classifying any generalized least-squares regression. In the theory, every symmetric least-squares regression begins from a generating function denoted by $\psi(x, y)$. The generating function is a positive, even, non-decreasing and homogenous function of the x and y deviations. The x and y deviations are then related to the slope of the line and a weight function $g(b) = \psi\left(1, \frac{1}{b}\right)$ is extracted from the error expression. In this way it is shown that every generalized symmetric least-squares problem is equivalent to a weighted ordinary least-squares problem, since the generalized error function is a product of $g(b)$ and Ehrenberg's ordinary least-squares error formula. All cases of symmetric regression are classified in terms of a specific generating function $\psi(x, y)$ and a corresponding weight function $g(b)$.

Even when a weight function $g(b)$ does not stem from a symmetric least-squares regression, as was the case with the hybrid symmetric regressions and the exponential regressions, one can still solve the weighted ordinary least-squares problem. In all cases, setting $E_a = 0$ and $E_b = 0$ leads to an implicit formula and an explicit formula for the discrepancy between the ordinary least-squares slope and the generalized least-squares slope, $b - b_{\text{OLS}}$. These formulas for b are called discrepancy formulas. The formula for a in all cases is given by $a = \mu_y - b\mu_x$.

The general calculation of the Hessian matrix and determinant produces a particular combination of $g(b)$ and its first and second derivatives. This expression is called the indicative function and denoted by $G(b)$. The indicative function streamlines the computation of the Hessian matrix and determinant. It subsequently also gives a simple way to test whether a regression arising from a weight function $g(b)$ has a minimizing solution: check whether $G(b)(b - b_{\text{OLS}}) > -1$. A table of indicative functions $G(b)$ for the specific regressions already described is presented.

Finally, the indicative equation (the differential equation expressing $G(b)$ in terms of $g(b)$) is solved, and a general integral formula expressing $g(b)$ in terms of $G(b)$ is obtained. The integral formula for $g(b)$ contains arbitrary constants, so that many different regressions actually have the same general weight function. Linear combinations of indicative functions are shown to produce valid weight functions. A table of indicative functions is presented with the corresponding integral weight functions worked out. The table reveals many disparate regressions belonging to the same general weight function class and having the same indicative function.

In this way, all symmetric least-squares regressions are categorized by a generating function, a weight function, and an indicative function. All weighted ordinary least-squares regressions, of which symmetric regressions are a part, are categorized by grouping them into classes with the same general indicative function and the same general weight function.

References

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

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

[3] 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.

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

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

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