

3-2015

Generalized Least-Squares Regressions V: Multiple Variables

Nataniel Greene
CUNY Kingsborough Community College

How does access to this work benefit you? Let us know!

Follow this and additional works at: http://academicworks.cuny.edu/kb_pubs

 Part of the [Mathematics Commons](#), [Numerical Analysis and Computation Commons](#), and the [Statistics and Probability Commons](#)

Recommended Citation

N. Greene. "Generalized Least-Squares Regressions V: Multiple Variables," in *New Developments in Pure and Applied Mathematics*, 2015, pp. 17-25.

This Article is brought to you for free and open access by the Kingsborough Community College at CUNY Academic Works. It has been accepted for inclusion in Publications and Research by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.

Generalized Least-Squares Regressions V: Multiple Variables

Nataniel Greene

Abstract—The multivariate theory of generalized least-squares is formulated here using the notion of generalized means. The multivariate generalized least-squares problem seeks an m dimensional hyperplane which minimizes the average generalized mean of the square deviations between the data and the hyperplane in $m + 1$ variables. The numerical examples presented suggest that a multivariate generalized least-squares method can be preferable to ordinary least-squares especially in situations where the data are ill-conditioned.

Keywords—Generalized least-squares, geometric mean regression, least-squares, multivariate regression, multiple regression, orthogonal regression.

I. OVERVIEW

ORDINARY least-squares regression in multiple variables x_0, x_1, \dots, x_m suffers from a fundamental lack of symmetry. It begins with the choice of one variable, x_0 , as the dependent variable and x_1, \dots, x_m as the independent variables. It then minimizes the distance between the data and the regression hyperplane in the x_0 variable alone. However, the regression hyperplane formed by minimizing the distance between the data and the hyperplane in the variable x_k is not the same as solving for x_k in the hyperplane formed by minimizing the distance in the variable x_j for $j \neq k$.

For each of the variables x_k , minimizing the distance between the data and the hyperplane in the variable x_k is called ordinary least-squares (OLS) $x_k | \{x_0, \dots, x_m\} \setminus \{x_k\}$ regression. To predict the value of x_k based on the data using OLS, one must use the OLS $x_k | \{x_0, \dots, x_m\} \setminus \{x_k\}$ regression hyperplane. It is not valid to take the OLS $x_j | \{x_0, \dots, x_m\} \setminus \{x_j\}$ hyperplane and solve for x_k when $j \neq k$.

The fact that there are $m + 1$ OLS hyperplanes to model a single set of data in $m + 1$ variables is problematic. One wishes to have a single linear model for the data, for which it is valid to solve for any one of the variables for prediction purposes. Multivariate generalized least-squares solves this problem by seeking to minimize the average generalized mean of the square deviations between the data and the hyperplane in all the variables simultaneously. For the resulting regression hyperplane, it is valid to solve for any of the variables for prediction purposes.

N. Greene is with the Department of Mathematics and Computer Science, Kingsborough Community College, City University of New York, 2001 Oriental Boulevard, Brooklyn, NY 11235 USA (phone: 718-368-5929; e-mail: ngreene.math@gmail.com).

II. MULTIVARIATE REGRESSIONS

The theory of generalized least-squares was already described by this author for the case of two variables [4]–[7]. The extension of this theory to multiple variables is now begun.

A. The Explicit Error Formula and Solution for Ordinary Least-Squares

The multivariate ordinary least-squares problem is defined as follows.

Definition 1: (Multivariate Ordinary Least-Squares Problem) An m dimensional hyperplane

$$x_0 = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m \quad (1)$$

is sought which minimizes the error function defined by

$$E = \frac{1}{N} \sum_{i=1}^N (\Delta x_{0i})^2 \quad (2)$$

where

$$\Delta x_{0i} = b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_mx_{mi} - x_{0i}. \quad (3)$$

This is called OLS $x_0 | \{x_1, \dots, x_m\}$ regression. In general, OLS $x_k | \{x_0, \dots, x_m\} \setminus \{x_k\}$ regression seeks a hyperplane which minimizes

$$E = \frac{1}{N} \sum_{i=1}^N (\Delta x_{ki})^2 \quad (4)$$

where

$$\Delta x_{ki} = \left(\frac{1}{b_k} x_{0i} - \frac{b_0}{b_k} - \frac{b_1}{b_k} x_{1i} - \dots - \frac{b_{k-1}}{b_k} x_{(k-1)i} - \frac{b_{k+1}}{b_k} x_{(k+1)i} - \dots - \frac{b_m}{b_k} x_{mi} \right) - x_{ki} \quad (5)$$

The deviation Δx_{ki} at the i th data point $(x_{0i}, x_{1i}, \dots, x_{mi})$ is the difference between the hyperplane solved in terms of the variable x_k and evaluated at the data point and the data value x_{ki} . Standard subscript notations are employed for dealing with means, standard deviations, correlation coefficients and covariances in multiple variables. The i th data value for the k th variable x_k is denoted by x_{ki} , which is short for $(x_k)_i$. The means, standard deviations, and correlation coefficients are denoted as follows: $\mu_k = \mu_{x_k}$, $\sigma_k = \sigma_{x_k}$, $\rho_{jk} = \rho_{x_j, x_k}$. The covariance notation $\sigma_{jk} = \rho_{jk} \sigma_j \sigma_k$ is preferred in this paper because it makes many of the complex formulas presented here more manageable. The notation $y = x_0$ and $\Delta y_i = \Delta x_{0i}$ can also be used. However, denoting the y -variable always using the zero subscript x_0 allows one to

easily identify it when there are any number of variables and naturally fits with the subscript convention just described.

The notation Δx_{ki} greatly simplifies working with the error function. The next lemma describes a fundamental relation between the regression coefficient b_k , Δx_{ki} , which is the deviation of the variable x_k from the hyperplane at the i th data value and Δx_{0i} , which is the deviation of x_0 from the hyperplane at the i th data value.

Lemma 2: (Fundamental Relation)

$$b_k = -\frac{\Delta x_{0i}}{\Delta x_{ki}} \quad (6)$$

or

$$\Delta x_{ki} = -\frac{1}{b_k} \cdot \Delta x_{0i} \quad (7)$$

The proof is straightforward algebra. This lemma will play a fundamental role further on in allowing one to always extract a weight function from any generalized least-squares error expression.

The explicit bivariate formula for the ordinary least-squares error described by Ehrenberg [3] has a known generalization using covariance notation. It is written here as a matrix-vector equation and also explicitly.

Theorem 3: (Explicit Multivariate Error Formula) Let $\mathbf{b} = (b_1, \dots, b_m)^T$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)^T$, $\mathbf{s}_0 = (\sigma_{10}, \dots, \sigma_{m0})^T$ and $\mathbf{S} = [\sigma_{jk}]_{m \times m}$. Then the multivariate ordinary least-squares error written in matrix-vector notation is

$$E = \mathbf{b}^T \mathbf{S} \mathbf{b} - 2\mathbf{b}^T \mathbf{s}_0 + \sigma_{00} + (b_0 - \mu_0 + \mathbf{b}^T \boldsymbol{\mu})^2 \quad (8)$$

and explicitly it is

$$E = \sum_{j=1}^m \sum_{k=1}^m \sigma_{jk} b_j b_k - 2 \sum_{k=1}^m \sigma_{k0} b_k \quad (9)$$

$$+ \sigma_{00} + \left(b_0 - \mu_0 + \sum_{k=1}^m b_k \mu_k \right)^2. \quad (10)$$

Alternatively write

$$E = \sum_{k=1}^m \sigma_{kk} b_k^2 + 2 \sum_{j < k} \sigma_{jk} b_j b_k - 2 \sum_{k=1}^m \sigma_{k0} b_k + \sigma_{00} + \left(b_0 - \mu_0 + \sum_{k=1}^m b_k \mu_k \right)^2. \quad (11)$$

Proof: Begin with the error expression and manipulate as follows.

$$\begin{aligned} E &= \frac{1}{N} \sum_{i=1}^N (b_0 + b_1 x_{1i} + \dots + b_m x_{mi} - x_{0i})^2 \\ &= \frac{1}{N} \sum_{i=1}^N (b_1 (x_{1i} - \mu_1) + \dots + b_m (x_{mi} - \mu_m) \\ &\quad - (x_{0i} - \mu_0) + (b_0 - \mu_0 + b_1 \mu_1 + \dots + b_m \mu_m))^2 \end{aligned}$$

Square the summand and distribute the summation onto each term. To simplify, utilize the covariance notation $\sigma_{jk} = \frac{1}{N} \sum_{i=1}^N (x_{ji} - \mu_j)(x_{ki} - \mu_k)$ and utilize the fact that

$\sum_{i=1}^N (x_{ji} - \mu_j) = 0$ for all $j = 0 \dots m$. The result is then obtained. ■

Corollary 4: The explicit error formula written in covariance notation for two variables x_1 and x_0 is

$$E = \sigma_{11} b_1^2 - 2\sigma_{10} b_1 + \sigma_{00} + (b_0 - \mu_0 + b_1 \mu_1)^2. \quad (12)$$

This is equivalent to Ehrenberg's formula. For three variables x_1, x_2 and x_0 , the formula is

$$E = \sigma_{11} b_1^2 + \sigma_{22} b_2^2 + 2\sigma_{12} b_1 b_2 - 2\sigma_{10} b_1 - 2\sigma_{20} b_2 + \sigma_{00} + (b_0 - \mu_0 + b_1 \mu_1 + b_2 \mu_2)^2 \quad (13)$$

For four variables x_1, x_2, x_3 and x_0 , the formula is

$$\begin{aligned} E &= \sigma_{11} b_1^2 + \sigma_{22} b_2^2 + \sigma_{33} b_3^2 \\ &\quad + 2\sigma_{12} b_1 b_2 + 2\sigma_{13} b_1 b_3 + 2\sigma_{23} b_2 b_3 \\ &\quad - 2\sigma_{10} b_1 - 2\sigma_{20} b_2 - 2\sigma_{30} b_3 \\ &\quad + \sigma_{00} + (b_0 - \mu_0 + b_1 \mu_1 + b_2 \mu_2 + b_3 \mu_3)^2 \end{aligned} \quad (14)$$

The explicit formula for the multivariate OLS regression coefficients is now written simply in matrix-vector form.

Theorem 5: (OLS explicit solution) The vector \mathbf{b} of OLS $x_0 | \{x_1, \dots, x_m\}$ regression coefficients is given explicitly by

$$\mathbf{b} = \mathbf{S}^{-1} \mathbf{s}_0 \quad (15)$$

and

$$b_0 = \mu_0 - \mathbf{b}^T \boldsymbol{\mu} \quad (16)$$

Proof: Let $\nabla = (\partial/\partial b_1, \dots, \partial/\partial b_m)^T$ denote the gradient operator. Take the gradient of the error with respect to \mathbf{b} and set it equal to zero: $\nabla E = \mathbf{0}$. Use the matrix-vector form of the error and distribute the gradient onto each term.

$$\nabla \left(\mathbf{b}^T \mathbf{S} \mathbf{b} - 2\mathbf{b}^T \mathbf{s}_0 + \sigma_{00} + (b_0 - \mu_0 + \mathbf{b}^T \boldsymbol{\mu})^2 \right) = \mathbf{0}$$

$$2\mathbf{S} \mathbf{b} - 2\mathbf{s}_0 - 2\boldsymbol{\mu} (b_0 - \mu_0 + \mathbf{b}^T \boldsymbol{\mu}) = \mathbf{0}$$

Set $b_0 = \mu_0 - \mathbf{b}^T \boldsymbol{\mu}$, and obtain

$$\mathbf{S} \mathbf{b} = \mathbf{s}_0$$

$$\mathbf{b} = \mathbf{S}^{-1} \mathbf{s}_0.$$

■

B. The Hessian Matrix

Since it is already known that the ordinary least-squares solution vector \mathbf{b} minimizes the error function, the Hessian matrix \mathbf{H} of second-order partial derivatives of E must be positive definite. Recall that \mathbf{H} is positive definite when $\det \mathbf{H} > 0$ and when the determinants of all the upper-left submatrices of \mathbf{H} are positive. Alternatively, \mathbf{H} is positive definite when all the eigenvalues of \mathbf{H} are positive. It is instructive here to compute the Hessian matrix.

Theorem 6: The Hessian matrix is given by

$$\mathbf{H} = 2 \begin{bmatrix} 1 & \mu_1 & \cdots & \mu_m \\ \mu_1 & \sigma_{11} + \mu_1^2 & \cdots & \sigma_{1m} + \mu_1 \mu_m \\ \vdots & \vdots & \ddots & \vdots \\ \mu_m & \sigma_{m1} + \mu_1 \mu_m & \cdots & \sigma_{mm} + \mu_m^2 \end{bmatrix}. \quad (17)$$

Proof: Form all second order partial derivatives of the error function

$$H_{(j+1)(k+1)} = \frac{\partial^2 E}{\partial b_k \partial b_j} \Big|_{(b_0, b_1, \dots, b_m)}$$

for $j, k = 0 \dots m$. Verify that

$$\begin{aligned} H_{11} &= 2 \\ H_{1(k+1)} &= H_{(k+1)1} = 2\mu_k \\ H_{(j+1)(k+1)} &= H_{(k+1)(j+1)} = 2(\sigma_{jk} + \mu_j \mu_k). \end{aligned}$$

Theorem 7: The Hessian determinant is given by

$$\det \mathbf{H} = 2^{m+1} \det \mathbf{S}$$

Proof: Taking a multiple of one row and adding it to another does not affect the determinant. For $k = 1 \dots m$, multiply the first row by $-\mu_k$ and add it to row $k + 1$. After this has been done to every row, perform a cofactor expansion along the first column and obtain the result.

$$\begin{aligned} \det \mathbf{H} &= 2^{m+1} \begin{vmatrix} 1 & \mu_1 & \cdots & \mu_m \\ \mu_1 & \sigma_{11} + \mu_1^2 & \cdots & \sigma_{1m} + \mu_1 \mu_m \\ \vdots & \vdots & \ddots & \vdots \\ \mu_m & \sigma_{m1} + \mu_1 \mu_m & \cdots & \sigma_{mm} + \mu_m^2 \end{vmatrix} \\ &= 2^{m+1} \begin{vmatrix} 1 & \mu_1 & \cdots & \mu_m \\ 0 & \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \sigma_{m1} & \cdots & \sigma_{mm} \end{vmatrix} \\ &= 2^{m+1} \begin{vmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{vmatrix} \end{aligned}$$

Theorem 8: The Hessian matrix \mathbf{H} and the covariance matrix \mathbf{S} are both positive definite.

Proof: Since $\det \mathbf{H} = 2^{m+1} \det \mathbf{S}$, the Hessian matrix \mathbf{H} is positive definite if and only if the covariance matrix \mathbf{S} is positive definite. However, it is already known that the OLS solution vector \mathbf{b} minimizes the error function. Therefore, conclude that \mathbf{H} and \mathbf{S} are both positive definite. ■

C. Generalized Means

Definition 9: A function $M(x_0, x_1, \dots, x_m)$ defines a generalized mean for all $x_i > 0$ if it satisfies Properties 1-5 below. If it satisfies Property 6 it is called a homogenous generalized mean. The properties are:

1. (Continuity) $M(x_0, x_1, \dots, x_m)$ is continuous in each variable.
2. (Monotonicity) $M(x_0, x_1, \dots, x_m)$ is non-decreasing in each variable.
3. (Symmetry)

$$M(x_0, x_1, \dots, x_m) = M(x_{s(0)}, x_{s(1)}, \dots, x_{s(m)})$$

where $s(i)$ is any permutation of the indices 0 through m .

4. (Identity)

$$M(x, x, \dots, x) = x.$$

5. (Intermediacy)

$$\min(x_0, \dots, x_m) \leq M(x_0, \dots, x_m) \leq \max(x_0, \dots, x_m)$$

6. (Homogeneity)

$$M(tx_0, tx_1, \dots, tx_m) = tM(x_0, x_1, \dots, x_m)$$

for all $t > 0$.

All the special multivariate means are included in this definition. Note that $m + 1$ variables are used in this definition in order that it share the same form as the $m + 1$ regression variables. XMR notation is used here to name generalized regressions: if 'X' is the letter used to denote a given generalized mean, then XMR is the corresponding generalized mean square regression.

Example 10: The multivariate harmonic mean is given by

$$H(x_0, \dots, x_m) = \frac{m + 1}{\frac{1}{x_0} + \dots + \frac{1}{x_m}}. \quad (18)$$

This mean generates multivariate orthogonal regression (HMR).

Example 11: The multivariate geometric mean is given by

$$G(x_0, \dots, x_m) = (x_0 \cdot \dots \cdot x_m)^{1/(m+1)}. \quad (19)$$

This mean generates multivariate geometric mean regression (GMR).

Example 12: The multivariate arithmetic mean is given by

$$A(x_0, \dots, x_m) = \frac{1}{m + 1} (x_0 + \dots + x_m). \quad (20)$$

This mean generates multivariate arithmetic mean regression (AMR).

Example 13: The selection mean is given by

$$S^{(k)}(x_0, \dots, x_m) = x_k \quad (21)$$

after x_0, \dots, x_m are arranged in increasing order. This mean generates OLS $x_k | \{x_0, \dots, x_m\} \setminus \{x_k\}$ regression.

Regressions based on these special cases are used in this work further on. The generalized means in the next examples have free parameters which can be used to parameterize these special cases.

Example 14: The power mean of order p is given by

$$M_p(x_0, \dots, x_m) = \left(\frac{1}{m + 1} (x_0^p + \dots + x_m^p) \right)^{1/p}. \quad (22)$$

Example 15: The weighted arithmetic mean with positive weights satisfying $\alpha_0 + \alpha_1 + \dots + \alpha_m = 1$, is given by

$$M_{(\alpha_0, \alpha_1, \dots, \alpha_m)}(x_0, x_1, \dots, x_m) = \alpha_0 x_0 + \alpha_1 x_1 + \dots + \alpha_m x_m \quad (23)$$

after x_0, \dots, x_m are arranged in increasing order.

Example 16: The weighted geometric mean with positive weights satisfying $\beta_0 + \beta_1 + \dots + \beta_m = 1$, is given by

$$M_{(\beta_0, \beta_1, \dots, \beta_m)}(x_0, x_1, \dots, x_m) = x_0^{\beta_0} x_1^{\beta_1} \dots x_m^{\beta_m} \quad (24)$$

after x_0, \dots, x_m are arranged in increasing order.

D. Two Generalized Least-Squares Problems and the Equivalence Theorem

The multivariate symmetric least-squares problem is formulated as follows.

Definition 17: (The Multivariate Symmetric Least-Squares Problem) An m dimensional hyperplane

$$x_0 = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m \quad (25)$$

is sought which minimizes the error function defined by

$$E = \frac{1}{N} \sum_{i=1}^N M \left((\Delta x_{0i})^2, (\Delta x_{1i})^2, \dots, (\Delta x_{mi})^2 \right) \quad (26)$$

where $M(x_0, x_1, \dots, x_m)$ is any generalized mean.

A more general related problem is the weighted ordinary least-squares problem.

Definition 18: (The Weighted Ordinary Least-Squares Problem) An m dimensional hyperplane

$$x_0 = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m \quad (27)$$

is sought which minimizes the error function defined by

$$E = g(b_1, \dots, b_m) \cdot \frac{1}{N} \sum_{i=1}^N (\Delta x_{0i})^2 \quad (28)$$

or simply $E = g \cdot E_{OLS}$.

The next theorem states that every multivariate symmetric least-squares problem is equivalent to a weighted multivariate ordinary least-squares problem with weight function $g(b_1, \dots, b_m)$.

Theorem 19: (Equivalence Theorem) Every general symmetric least-squares error function can be written equivalently as

$$E = g(b_1, \dots, b_m) \cdot \frac{1}{N} \sum_{i=1}^N (\Delta x_{0i})^2 \quad (29)$$

or simply $E = g \cdot E_{OLS}$ with

$$g(b_1, \dots, b_m) = M \left(1, \frac{1}{b_1^2}, \dots, \frac{1}{b_m^2} \right) \quad (30)$$

and E_{OLS} expressed using the explicit error formula.

Proof: Write

$$\begin{aligned} E &= \frac{1}{N} \sum_{i=1}^N M \left((\Delta x_{0i})^2, (\Delta x_{1i})^2, \dots, (\Delta x_{mi})^2 \right) \\ &= \frac{1}{N} \sum_{i=1}^N M \left((\Delta x_{0i})^2, \frac{1}{b_1^2} (\Delta x_{0i})^2, \dots, \frac{1}{b_m^2} (\Delta x_{0i})^2 \right) \\ &= \frac{1}{N} \sum_{i=1}^N (\Delta x_{0i})^2 M \left(1, \frac{1}{b_1^2}, \dots, \frac{1}{b_m^2} \right) \end{aligned}$$

where the fundamental relation

$$(\Delta x_{ki})^2 = \frac{1}{b_k^2} (\Delta x_{0i})^2$$

is used. Let $g(b_1, \dots, b_m) = M \left(1, \frac{1}{b_1^2}, \dots, \frac{1}{b_m^2} \right)$ and factor it out from the summation. ■

Example 20: The weight function corresponding to HMR is given by

$$g(b_1, \dots, b_m) = \frac{m+1}{1+b_1^2+\dots+b_m^2}. \quad (31)$$

Example 21: The weight function corresponding to GMR is given by

$$g(b_1, \dots, b_m) = (b_1 \cdot \dots \cdot b_m)^{-2/(m+1)}. \quad (32)$$

Example 22: The weight function corresponding to AMR is given by

$$g(b_1, \dots, b_m) = \frac{1}{m+1} \left(1 + \frac{1}{b_1^2} + \dots + \frac{1}{b_m^2} \right). \quad (33)$$

Example 23: The weight function corresponding to the k th selection mean is given by

$$g(b_1, \dots, b_m) = \frac{1}{b_k^2}. \quad (34)$$

Example 24: The weight function corresponding to the power mean is given by

$$g_p(b_1, \dots, b_m) = \left(\frac{1}{m+1} \left(1 + b_1^{-2p} + \dots + b_m^{-2p} \right) \right)^{1/p}. \quad (35)$$

Example 25: The weight function corresponding to the weighted arithmetic mean is

$$g(b_1, \dots, b_m) = \alpha_0 + \alpha_1 b_1^{-2} + \dots + \alpha_m b_m^{-2}. \quad (36)$$

Example 26: The weight function corresponding to the weighted geometric mean is given by

$$g(b_1, \dots, b_m) = b_1^{-2\beta_1} \cdot \dots \cdot b_m^{-2\beta_m}. \quad (37)$$

E. Solving for the Generalized Regression Coefficients

The fundamental practical question of multivariate generalized regression is how to solve for the coefficients b_1, \dots, b_m . The next theorem describes the procedure in general. The procedure is applied further on to produce the specific regression equations in three and four variables for several cases of interest.

Theorem 27: (System of Equations for Generalized Regression Coefficients) Let E denote the ordinary least-squares error function, $g_k = \partial g / \partial b_k$, $\nabla g = (g_1, \dots, g_m)^T$ and

$$F = \mathbf{b}^T \mathbf{S} \mathbf{b} - 2\mathbf{b}^T \mathbf{s}_0 + \sigma_{00}. \quad (38)$$

Then the vector $\mathbf{b} = (b_1, \dots, b_m)^T$ of regression coefficients is obtained by solving the nonlinear matrix-vector equation

$$F \nabla g + 2g(\mathbf{S} \mathbf{b} - \mathbf{s}_0) = \mathbf{0} \quad (39)$$

for \mathbf{b} and $b_0 = \mu_0 - \mathbf{b}^T \boldsymbol{\mu}$. Explicitly, one solves the nonlinear system

$$g_k F + g F_k = 0 \quad (40)$$

for b_1, \dots, b_m where $k = 1 \dots m$,

$$F = \sum_{j=1}^m \sum_{k=1}^m \sigma_{jk} b_j b_k - 2 \sum_{k=1}^m \sigma_{k0} b_k + \sigma_{00} \quad (41)$$

and $F_k = \partial F / \partial b_k$ is given by

$$F_k = 2 \sum_{j=1}^m \sigma_{jk} b_j - 2\sigma_{k0}. \quad (42)$$

Proof: Let $E_g = gE$ be the generalized regression error where $E = E_{OLS}$. Use the matrix-vector form of the error, take the gradient of the error with respect to \mathbf{b} and set it equal to zero.

$$\begin{aligned}\nabla(gE) &= \mathbf{0} \\ E\nabla g + g\nabla E &= \mathbf{0}\end{aligned}$$

Substitute

$$E = \mathbf{b}^T \mathbf{S} \mathbf{b} - 2\mathbf{b}^T \mathbf{s}_0 + \sigma_{00} + (b_0 - \mu_0 + \mathbf{b}^T \boldsymbol{\mu})^2$$

and

$$\nabla E = 2(\mathbf{S} \mathbf{b} - \mathbf{s}_0) - 2\boldsymbol{\mu}(b_0 - \mu_0 + \mathbf{b}^T \boldsymbol{\mu}).$$

Set $b_0 = \mu_0 - \mathbf{b}^T \boldsymbol{\mu}$, and obtain

$$2g(\mathbf{S} \mathbf{b} - \mathbf{s}_0) + F\nabla g = \mathbf{0} \quad (43)$$

Alternatively, take the partial derivative of E_g with respect to b_k for $k = 1 \dots m$ and set the resulting expressions equal to zero. ■

F. The Hessian Matrix and Determinant

In order for the regression coefficients b_0, \dots, b_m to minimize the error function and be admissible, the Hessian matrix of second order partial derivatives must be positive definite when evaluated at b_0, \dots, b_m . The general Hessian matrix is calculated next. As in the bivariate case, certain combinations of g and its first and second partial derivatives appear in the matrix. One combination is denoted here by J_{jk} and another is denoted by G_{jk} . They are called indicative functions.

Definition 28: Define the indicative functions

$$J_{jk} = \frac{g_{jk}}{g} - \frac{2g_j g_k}{g^2} \quad (44)$$

and

$$G_{jk} = \frac{2g_j}{g} - \frac{g_{jk}}{g_k}. \quad (45)$$

The two indicative functions are related by the equation $G_{jk} F_k = J_{jk} F$.

Theorem 29: (Hessian matrix) The Hessian matrix \mathbf{H} of second order partial derivatives of the error function given by

$$H_{(j+1)(k+1)} = \left. \frac{\partial^2}{\partial b_k \partial b_j} (gE) \right|_{(b_0, b_1, \dots, b_m)} \quad (46)$$

for $j, k = 0 \dots m$. It is computed explicitly as follows.

$$H_{11} = 2g \quad (47)$$

$$H_{1(k+1)} = H_{(k+1)1} = 2g\mu_k \quad (48)$$

$$H_{(j+1)(k+1)} = H_{(k+1)(j+1)} = g(J_{jk} F + 2\sigma_{jk} + 2\mu_j \mu_k) \quad (49)$$

Alternatively,

$$H_{(j+1)(k+1)} = g(G_{jk} F_k + 2\sigma_{jk} + 2\mu_j \mu_k). \quad (50)$$

Proof: Take the second order partial derivative

$$\begin{aligned}\frac{\partial^2}{\partial b_k \partial b_j} (gE) &= \frac{\partial}{\partial b_k} (g_j E + g E_j) \\ &= g_{jk} E + g_j E_k + g_k E_j + g E_{jk}.\end{aligned}$$

Since

$$g_k E + g E_k = 0$$

substitute $E_k = -\frac{g_k}{g} E$ and $E_j = -\frac{g_j}{g} E$ into the two middle terms, simplify, and obtain

$$\frac{\partial^2}{\partial b_k \partial b_j} (gE) = g \left(\left(\frac{g_{jk}}{g} - \frac{2g_j g_k}{g^2} \right) E + E_{jk} \right)$$

which is the first form of the Hessian. Now substitute $E = -\frac{g}{g_k} E_k$ and obtain the second form

$$\frac{\partial^2}{\partial b_k \partial b_j} (gE) = g \left(\left(\frac{2g_j}{g} - \frac{g_{jk}}{g_k} \right) E_k + E_{jk} \right).$$

As before, upon substituting for b_0 , $E = F$, $E_k = F_k$ and $E_{jk} = F_{jk} = 2\sigma_{jk} + 2\mu_j \mu_k$. ■

Theorem 30: (Hessian determinant) The Hessian determinant is given by

$$\det \mathbf{H} = g^{m+1} \det (F\mathbf{J} + 2\mathbf{S}) \quad (51)$$

where $\mathbf{J} = [J_{jk}]_{m \times m}$ and explicitly by

$$\det \mathbf{H} = g^{m+1} \begin{vmatrix} FJ_{11} + 2\sigma_{11} & \cdots & FJ_{m1} + 2\sigma_{1m} \\ \vdots & \ddots & \vdots \\ FJ_{m1} + 2\sigma_{1m} & \cdots & FJ_{mm} + 2\sigma_{mm} \end{vmatrix}. \quad (52)$$

Alternatively,

$$\det \mathbf{H} = g^{m+1} \det (\mathbf{K} + 2\mathbf{S}) \quad (53)$$

where $\mathbf{K} = [G_{jk} F_k]_{m \times m}$ and explicitly by

$$\det \mathbf{H} = g^{m+1} \begin{vmatrix} G_{11} F_1 + 2\sigma_{11} & \cdots & G_{1m} F_m + 2\sigma_{1m} \\ \vdots & \ddots & \vdots \\ G_{m1} F_1 + 2\sigma_{1m} & \cdots & G_{mm} F_m + 2\sigma_{mm} \end{vmatrix}. \quad (54)$$

Proof: Begin with the $(m+1) \times (m+1)$ determinant of \mathbf{H} and reduce it to an equivalent $m \times m$ determinant. The $(m+1) \times (m+1)$ determinant is given by

$$\det \mathbf{H} = g^{m+1} \begin{vmatrix} 2 & & 2\mu_1 \\ 2\mu_1 & G_{11} F_1 + 2\sigma_{11} + 2\mu_1^2 & \\ \vdots & \vdots & \\ 2\mu_m & G_{1m} F_1 + 2\sigma_{1m} + 2\mu_1 \mu_m & \\ \cdots & 2\mu_m & \\ \cdots & G_{1m} F_m + 2\sigma_{1m} + 2\mu_1 \mu_m & \\ \vdots & \vdots & \\ \cdots & G_{1m} F_m + 2\sigma_{mm} + 2\mu_m^2 & \end{vmatrix} \quad (55)$$

or alternatively

$$\det \mathbf{H} = g^{m+1} \begin{vmatrix} 2 & & & & 2\mu_1 \\ 2\mu_1 & J_{11}F + 2\sigma_{11} + 2\mu_1^2 & & & \\ \vdots & & \ddots & & \\ 2\mu_m & J_{1m}F + 2\sigma_{1m} + 2\mu_1\mu_m & & & \\ \dots & & & & 2\mu_m \\ \dots & J_{1m}F + 2\sigma_{1m} + 2\mu_1\mu_m & & & \\ \vdots & & \ddots & & \\ \dots & J_{1m}F + 2\sigma_{mm} + 2\mu_m^2 & & & \end{vmatrix}. \quad (56)$$

As was done for the OLS Hessian, multiply the first row by $-\mu_k$ and add it to row $k + 1$ for $k = 1 \dots m$. After this has been done to every row, perform a cofactor expansion along the first column and obtain the result. ■

G. Least-Volume and Hybrid Least-Volume Regression

A multivariate regression method is described by Tofallis [9]–[11] which minimizes the hypervolumes of the simplices formed by the data and the regression hyperplane. It is referred to here as Tofallis' least-volume regression (LVR).

Definition 31: (LVR) Let $V = x_0 \cdot x_1 \cdot \dots \cdot x_m$ denote the hypervolume of dimension $m + 1$ for all $x_k > 0$. An m dimensional hyperplane $x_0 = b_0 + b_1x_1 + \dots + b_mx_m$ is sought which minimizes the average hypervolume formed by the data points and the hyperplane. The error function is given by

$$E = \frac{1}{N} \sum_{i=1}^N V(\Delta x_{0i}, \Delta x_{1i}, \dots, \Delta x_{mi}). \quad (57)$$

For the case of bivariate regression ($m = 1$), least-volume regression is the same as geometric mean regression. Multivariate GMR, first described in a paper by Draper and Yang [2], is equivalent to minimizing the average of the hypervolumes raised to the $2/(m + 1)$ power. For $m \neq 1$, LVR is not equivalent to GMR and it is not a generalized least-squares regression. It is instead a generalized least- p th power regression with $p = m + 1$.

Theorem 32: Least-volume regression in $m + 1$ variables is equivalent to a weighted least- p th power regression with $p = m + 1$.

Proof: Write

$$\begin{aligned} E &= \frac{1}{N} \sum_{i=1}^N V(|\Delta x_{0i}|, |\Delta x_{1i}|, \dots, |\Delta x_{mi}|) \\ &= \frac{1}{N} \sum_{i=1}^N V\left(|\Delta x_{0i}|, \frac{1}{|b_1|} |\Delta x_{0i}|, \dots, \frac{1}{|b_m|} |\Delta x_{0i}|\right) \\ &= \frac{1}{N} \sum_{i=1}^N |\Delta x_{0i}|^{m+1} V\left(1, \frac{1}{|b_1|}, \dots, \frac{1}{|b_m|}\right) \\ &= \left(\frac{1}{|b_1 \cdot \dots \cdot b_m|}\right) \frac{1}{N} \sum_{i=1}^N |\Delta x_{0i}|^{m+1} \end{aligned}$$

The weight function for LVR is given by

$$g(b_1, \dots, b_m) = 1/|b_1 \cdot \dots \cdot b_m|. \quad \blacksquare$$

A related weighted least-squares method is now defined giving close results to LVR called hybrid least-volume regression. It is a hybrid regression method along the lines of the hybrid methods discussed in the first paper of this series [4].

Definition 33: (Hybrid LVR) An m dimensional hyperplane is sought which minimizes the hybrid error function

$$E = g(b_1, \dots, b_m) \cdot \frac{1}{N} \sum_{i=1}^N (\Delta x_{0i})^2 \quad (58)$$

where

$$g(b_1, \dots, b_m) = \frac{1}{|b_1 \cdot \dots \cdot b_m|}. \quad (59)$$

The error function is a product of the LVR weight function and the multivariate OLS error function.

The hybrid LVR coefficients obtained in the second example below are seen to differ from the actual LVR coefficients only in the hundredths place.

H. Specific Regression Equations for the Case of Three Variables

The general formula for the regression coefficients is applied here to the problem of determining the coefficients in the equation

$$x_0 = b_0 + b_1x_1 + b_2x_2 \quad (60)$$

for certain special cases. Again, covariance notation σ_{jk} is used in order to obtain equations that are as simple as possible to write. In all cases, $b_0 = \mu_0 - b_1\mu_1 - b_2\mu_2$.

For OLS $x_0\{x_1, x_2\}$, which is standard OLS regression corresponding to the selection mean $S^{(0)}(x_0, x_1, x_2)$ and a weight function $g(b_1, b_2) = 1$, the following linear system of equations in b_1 and b_2 is obtained.

$$\begin{cases} \sigma_{11}b_1 + \sigma_{12}b_2 = \sigma_{10} \\ \sigma_{12}b_1 + \sigma_{22}b_2 = \sigma_{20} \end{cases} \quad (61)$$

For OLS $x_1\{x_2, x_0\}$ corresponding to the selection mean $S^{(1)}(x_0, x_1, x_2)$ and the weight function $g(b_1, b_2) = \frac{1}{b_1^2}$, the following system of equations is obtained.

$$\begin{cases} \sigma_{22}b_2^2 + \sigma_{12}b_1b_2 - \sigma_{10}b_1 - 2\sigma_{20}b_2 + \sigma_{00} = 0 \\ \sigma_{12}b_1 + \sigma_{22}b_2 - \sigma_{20} = 0 \end{cases} \quad (62)$$

For OLS $x_2\{x_1, x_0\}$ corresponding to the selection mean $S^{(2)}(x_0, x_1, x_2)$ and the weight function $g(b_1, b_2) = \frac{1}{b_2^2}$, the following system of equations is obtained.

$$\begin{cases} \sigma_{11}b_1 + \sigma_{12}b_2 - \sigma_{10} = 0 \\ \sigma_{11}b_1^2 + \sigma_{12}b_1b_2 - 2\sigma_{10}b_1 - \sigma_{20}b_2 + \sigma_{00} = 0 \end{cases} \quad (63)$$

For HMR, which is orthogonal regression, the following system of equations is obtained.

$$\begin{cases} \sigma_{12}b_2^3 + (\sigma_{11} - \sigma_{22})b_1b_2^2 - \sigma_{12}b_1^2b_2 + \sigma_{10}b_1^2 + 2\sigma_{20}b_1b_2 \\ \quad - \sigma_{10}b_2^2 - (\sigma_{00} - \sigma_{11})b_1 + \sigma_{12}b_2 - \sigma_{10} = 0 \\ \sigma_{12}b_1^3 - (\sigma_{11} - \sigma_{22})b_1^2b_2 - \sigma_{12}b_1b_2^2 + \sigma_{20}b_2^2 + \sigma_{10}b_1b_2 \\ \quad - \sigma_{20}b_1^2 - (\sigma_{00} - \sigma_{22})b_2 + \sigma_{12}b_1 - \sigma_{20} = 0 \end{cases} \quad (64)$$

For GMR the following system of equations is obtained.

$$\begin{cases} 2\sigma_{11}b_1^2 - \sigma_{22}b_2^2 + \sigma_{12}b_1b_2 - \sigma_{10}b_1 + 2\sigma_{20}b_2 - \sigma_{00} = 0 \\ \sigma_{11}b_1^2 - 2\sigma_{22}b_2^2 - \sigma_{12}b_1b_2 - 2\sigma_{10}b_1 + \sigma_{20}b_2 + \sigma_{00} = 0 \end{cases} \quad (65)$$

For AMR the following system of equations is obtained.

$$\begin{cases} \sigma_{11}b_1^4b_2^2 + \sigma_{12}b_1^3b_2^3 - \sigma_{10}b_1^3b_2^2 + \sigma_{11}b_1^4 - \sigma_{22}b_2^4 + \sigma_{12}b_1^3b_2 \\ -\sigma_{12}b_2^3b_1 + 2\sigma_{20}b_2^2 - \sigma_{10}b_1^3 + \sigma_{10}b_2^2b_1 - \sigma_{00}b_2^2 = 0 \\ \sigma_{22}b_1^2b_2^4 + \sigma_{12}b_1^3b_2^3 - \sigma_{20}b_1^2b_2^2 - \sigma_{11}b_1^4 + \sigma_{22}b_2^4 + \sigma_{12}b_1b_2^3 \\ -\sigma_{12}b_1^3b_2 + 2\sigma_{10}b_1^3 - \sigma_{20}b_2^2 + \sigma_{20}b_1^2b_2 - \sigma_{00}b_1^2 = 0 \end{cases} \quad (66)$$

For hybrid LVR the following system of equations is obtained.

$$\begin{cases} \sigma_{11}b_1^2 - \sigma_{22}b_2^2 + 2\sigma_{20}b_2 - \sigma_{00} = 0 \\ \sigma_{11}b_1^2 - \sigma_{22}b_2^2 - 2\sigma_{10}b_1 + \sigma_{00} = 0 \end{cases} \quad (67)$$

I. Specific Regression Equations for the Case of Four Variables

The general formula for the regression coefficients is applied here to the problem of determining the coefficients in the equation

$$x_0 = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \quad (68)$$

for certain special cases. In all cases, $b_0 = \mu_0 - b_1\mu_1 - b_2\mu_2 - b_3\mu_3$.

For OLS $x_0 | \{x_1, x_2, x_3\}$ regression, which is standard ordinary least-squares, the following system of equations is obtained.

$$\begin{cases} \sigma_{11}b_1 + \sigma_{12}b_2 + \sigma_{13}b_3 = \sigma_{10} \\ \sigma_{12}b_1 + \sigma_{22}b_2 + \sigma_{23}b_3 = \sigma_{20} \\ \sigma_{13}b_1 + \sigma_{23}b_2 + \sigma_{33}b_3 = \sigma_{30} \end{cases} \quad (69)$$

For OLS $x_1 | \{x_2, x_3, x_0\}$ regression, the following system of equations is obtained.

$$\begin{cases} \sigma_{22}b_2^2 + \sigma_{33}b_3^2 + \sigma_{12}b_1b_2 + 2\sigma_{23}b_2b_3 + \sigma_{13}b_1b_3 \\ -\sigma_{10}b_1 - 2\sigma_{20}b_2 - 2\sigma_{30}b_3 + \sigma_{00} = 0 \\ \sigma_{12}b_1 + \sigma_{22}b_2 + \sigma_{23}b_3 - \sigma_{20} = 0 \\ \sigma_{13}b_1 + \sigma_{23}b_2 + \sigma_{33}b_3 - \sigma_{30} = 0 \end{cases} \quad (70)$$

For OLS $x_2 | \{x_1, x_3, x_0\}$ regression, the following system of equations is obtained.

$$\begin{cases} \sigma_{11}b_1 + \sigma_{12}b_2 + \sigma_{13}b_3 - \sigma_{10} = 0 \\ \sigma_{11}b_1^2 + \sigma_{33}b_3^2 + \sigma_{12}b_1b_2 + \sigma_{23}b_2b_3 + 2\sigma_{13}b_1b_3 \\ -2\sigma_{10}b_1 - \sigma_{20}b_2 - 2\sigma_{30}b_3 + \sigma_{00} = 0 \\ \sigma_{13}b_1 + \sigma_{23}b_2 + \sigma_{33}b_3 - \sigma_{30} = 0 \end{cases} \quad (71)$$

For OLS $x_3 | \{x_1, x_2, x_0\}$ regression, the following system of equations is obtained.

$$\begin{cases} \sigma_{11}b_1 + \sigma_{12}b_2 + \sigma_{13}b_3 - \sigma_{10} = 0 \\ \sigma_{12}b_1 + \sigma_{22}b_2 + \sigma_{23}b_3 - \sigma_{20} = 0 \\ \sigma_{11}b_1^2 + \sigma_{22}b_2^2 + 2\sigma_{12}b_1b_2 + \sigma_{23}b_2b_3 + \sigma_{13}b_1b_3 \\ -2\sigma_{10}b_1 - 2\sigma_{20}b_2 - \sigma_{30}b_3 + \sigma_{00} = 0 \end{cases} \quad (72)$$

For HMR the following system of equations is obtained.

$$\begin{cases} (\sigma_{22} - \sigma_{11})b_1b_2^2 + \sigma_{12}b_1^2b_2 + \sigma_{33}b_1b_2^3 - \sigma_{13}b_2^2b_3 \\ -\sigma_{11}b_1b_2^3 + \sigma_{13}b_1^2b_3 - \sigma_{12}b_2b_2^3 + 2\sigma_{23}b_1b_2b_3 \\ -\sigma_{12}b_2^3 - \sigma_{13}b_3^3 - \sigma_{10}b_1^2 + \sigma_{10}b_2^2 \\ +\sigma_{10}b_2^3 - 2\sigma_{20}b_1b_2 - 2\sigma_{30}b_1b_3 + (\sigma_{00} - \sigma_{11})b_1 \\ -\sigma_{12}b_2 - \sigma_{13}b_3 + \sigma_{10} = 0 \\ -\sigma_{12}b_1b_2^2 - \sigma_{33}b_2b_2^3 - 2\sigma_{13}b_1b_2b_3 - \sigma_{23}b_2^2b_3 \\ + (\sigma_{22} - \sigma_{11})b_1^2b_2 + \sigma_{22}b_2b_2^3 + \sigma_{12}b_1b_2^3 + \sigma_{23}b_1^2b_3 \\ +\sigma_{12}b_1^3 + \sigma_{23}b_3^3 - \sigma_{20}b_1^2 + \sigma_{20}b_2^2 \\ -\sigma_{20}b_2^3 + 2\sigma_{30}b_2b_3 + 2\sigma_{10}b_1b_2 + \sigma_{12}b_1 \\ + (\sigma_{22} - \sigma_{00})b_2 + \sigma_{23}b_3 - \sigma_{20} = 0 \\ -\sigma_{13}b_1b_2^3 + (\sigma_{33} - \sigma_{11})b_1^2b_3 - \sigma_{23}b_2b_2^3 - \sigma_{22}b_2^2b_3 \\ -2\sigma_{12}b_1b_2b_3 + \sigma_{23}b_3^3 + \sigma_{13}b_1^3 + \sigma_{33}b_2^2b_3 \\ +\sigma_{23}b_1^2b_2 + \sigma_{13}b_1b_2^2 - \sigma_{30}b_1^2 - \sigma_{30}b_2^2 \\ +\sigma_{30}b_2^3 + 2\sigma_{20}b_2b_3 + 2\sigma_{10}b_1b_3 \\ +\sigma_{13}b_1 + \sigma_{23}b_2 + (\sigma_{33} - \sigma_{00})b_3 - \sigma_{30} = 0 \end{cases} \quad (73)$$

For GMR, the following system of equations is obtained.

$$\begin{cases} 3\sigma_{11}b_1^2 - \sigma_{22}b_2^2 - \sigma_{33}b_3^2 + 2\sigma_{12}b_1b_2 - 2\sigma_{23}b_2b_3 \\ +2\sigma_{13}b_1b_3 - 2\sigma_{10}b_1 + 2\sigma_{20}b_2 + 2\sigma_{30}b_3 - \sigma_{00} = 0 \\ \sigma_{11}b_1^2 - 3\sigma_{22}b_2^2 + \sigma_{33}b_3^2 - 2\sigma_{12}b_1b_2 - 2\sigma_{23}b_2b_3 \\ +2\sigma_{13}b_1b_3 - 2\sigma_{10}b_1 + 2\sigma_{20}b_2 - 2\sigma_{30}b_3 + \sigma_{00} = 0 \\ \sigma_{11}b_1^2 + \sigma_{22}b_2^2 - 3\sigma_{33}b_3^2 + 2\sigma_{12}b_1b_2 - 2\sigma_{23}b_2b_3 \\ -2\sigma_{13}b_1b_3 - 2\sigma_{10}b_1 - 2\sigma_{20}b_2 + 2\sigma_{30}b_3 + \sigma_{00} = 0 \end{cases} \quad (74)$$

For AMR the following system of equations is obtained.

$$\begin{cases} \sigma_{11}b_1^4b_2^2b_3^2 + \sigma_{12}b_1^3b_2^3b_3^2 + \sigma_{13}b_1^3b_2^2b_3^3 - \sigma_{10}b_1^3b_2^2b_3^2 \\ +\sigma_{12}b_1^2b_2b_2^3 + \sigma_{13}b_1^2b_2^2b_3 - \sigma_{12}b_1b_2^2b_3^2 - \sigma_{13}b_1b_2^2b_3^3 \\ \sigma_{11}b_1^4b_2^3 - \sigma_{22}b_2^4b_3^2 - \sigma_{33}b_2^4b_3^3 - 2\sigma_{23}b_2^3b_3^3 + \sigma_{13}b_1^3b_3^3 \\ +\sigma_{11}b_1^4b_2^2 + \sigma_{12}b_1^3b_2^3 + \sigma_{10}b_1b_2^2b_3^2 + 2\sigma_{20}b_2^2b_3^2 \\ +2\sigma_{30}b_2^2b_3^3 - \sigma_{10}b_1^3b_2^2 - \sigma_{10}b_1^3b_2^2 - \sigma_{00}b_2^2b_3^2 = 0 \\ \sigma_{22}b_1^2b_2^4b_3^2 + \sigma_{23}b_1^2b_2^3b_3^3 + \sigma_{12}b_1^2b_2^3b_3^2 - \sigma_{20}b_1^2b_2^3b_3^2 \\ -\sigma_{12}b_1^2b_2b_2^3 - \sigma_{23}b_1^2b_2b_3^3 + \sigma_{23}b_1^2b_2^2b_3 + \sigma_{12}b_1b_2^3b_3^2 \\ -\sigma_{11}b_1^4b_2^3 + \sigma_{22}b_2^4b_3^2 + \sigma_{23}b_2^3b_3^3 + \sigma_{12}b_1^2b_2^3 - 2\sigma_{13}b_1^3b_3^3 \\ -\sigma_{33}b_1^2b_3^4 + \sigma_{22}b_1^2b_2^4 + \sigma_{20}b_1^2b_2b_2^3 - \sigma_{20}b_3^3b_3^2 + 2\sigma_{10}b_1^3b_2^3 \\ +2\sigma_{30}b_1^2b_3^3 - \sigma_{20}b_1^2b_3^2 - \sigma_{00}b_1^2b_2^3 = 0 \\ \sigma_{13}b_1^3b_2^2b_3^3 + \sigma_{23}b_1^2b_2^3b_3^3 + \sigma_{33}b_1^2b_2^2b_3^3 - \sigma_{30}b_1^2b_2^2b_3^3 \\ -\sigma_{13}b_1^2b_2^2b_3 + \sigma_{13}b_1b_2^2b_3^3 - \sigma_{23}b_1^2b_2^2b_3 + \sigma_{23}b_1^2b_2b_3^3 \\ +\sigma_{33}b_2^2b_3^4 + \sigma_{23}b_2^3b_3^3 + \sigma_{13}b_1^3b_3^3 - \sigma_{11}b_1^4b_2^2 - 2\sigma_{12}b_1^2b_2^3 \\ +\sigma_{33}b_1^2b_3^4 - \sigma_{22}b_1^2b_2^4 - \sigma_{30}b_2^2b_3^3 + 2\sigma_{10}b_1^3b_2^2 - \sigma_{30}b_1^2b_3^3 \\ +2\sigma_{20}b_1^2b_2^3 + \sigma_{30}b_1^2b_2^2b_3 - \sigma_{00}b_1^2b_2^2 = 0 \end{cases} \quad (75)$$

For hybrid LVR the following system of equations is obtained.

$$\begin{cases} \sigma_{11}b_1^2 - \sigma_{22}b_2^2 - \sigma_{33}b_3^2 - 2\sigma_{23}b_2b_3 \\ +2\sigma_{20}b_2 + 2\sigma_{30}b_3 - \sigma_{00} = 0 \\ \sigma_{11}b_1^2 - \sigma_{22}b_2^2 + \sigma_{33}b_3^2 + 2\sigma_{13}b_1b_3 \\ -2\sigma_{10}b_1 - 2\sigma_{30}b_3 + \sigma_{00} = 0 \\ \sigma_{11}b_1^2 + \sigma_{22}b_2^2 - \sigma_{33}b_3^2 + 2\sigma_{12}b_1b_2 \\ -2\sigma_{10}b_1 - 2\sigma_{20}b_2 + \sigma_{00} = 0 \end{cases} \quad (76)$$

III. NUMERICAL EXAMPLES

Example 34: (Cement Data) This example is taken from Hald's statistics text [8] (p. 636). The data describe the heat evolved in the curing of cement as a function of the percentage in weight of certain compounds in the mixture. The y variable, called here x_0 , is the heat measured in calories per gram.

The variables x_1 and x_2 are the percentages by weight of two different cement compounds. The data are as follows.

x_1	7	1	11	11	7	11	3	1	2	21	1	11	10
x_2	26	29	56	31	52	55	71	31	54	47	40	66	68
x_0	78.5	74.3	104.3	87.6	95.9	109.2	102.7	72.5	93.1	115.9	83.8	113.3	109.4

The reader can verify that $\mu_0 = 95.42308$, $\sigma_{00} = 208.90485$,

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 7.46154 \\ 48.15385 \end{bmatrix}, \quad (77)$$

$$\mathbf{s}_0 = \begin{bmatrix} \sigma_{10} \\ \sigma_{20} \end{bmatrix} = \begin{bmatrix} 59.68935 \\ 176.38106 \end{bmatrix}, \quad (78)$$

and

$$\mathbf{S} = \begin{bmatrix} 31.94082 & 19.31361 \\ 19.31360 & 223.51479 \end{bmatrix}. \quad (79)$$

The data are well-conditioned: $\text{cond } \mathbf{S} = 7.51147$. Also $\det \mathbf{S} = 6.76623 \times 10^3$. The standard OLS $x_0 | \{x_1, x_2\}$ regression plane is given by

$$x_0 = 52.5773 + 1.4683x_1 + 0.6623x_2. \quad (80)$$

The OLS $x_1 | \{x_2, x_0\}$ regression plane is given by

$$x_0 = 52.2466 + 1.5685x_1 + 0.6536x_2. \quad (81)$$

The OLS $x_2 | \{x_1, x_0\}$ regression plane is given by

$$x_0 = 51.1917 + 1.4491x_1 + 0.6940x_2. \quad (82)$$

The HMR plane is given by

$$x_0 = 52.2110 + 1.5271x_1 + 0.6607x_2. \quad (83)$$

The GMR plane is given by

$$x_0 = 52.0069 + 1.4950x_1 + 0.6700x_2. \quad (84)$$

The AMR plane is given by

$$x_0 = 51.7164 + 1.4699x_1 + 0.6799x_2. \quad (85)$$

The hybrid LVR plane is given by

$$x_0 = 51.7166 + 1.5085x_1 + 0.6739x_2. \quad (86)$$

This example suggests that all the methods yield regression planes that are reasonably close to each other when the data are well-conditioned. The next example illustrates that when the data are ill-conditioned, ordinary least-squares can perform poorly while generalized least-squares methods can still perform well.

Example 35: This example is chosen from the work of Tofallis [10], [11] where the data are used to compare least-volume regression to ordinary least-squares. The data are from a model problem in Belsley's book on collinearity [1] (p. 5) and are ill-conditioned.

Suppose an underlying linear model of some physical relationship is known and given by

$$x_0 = 1.2 - 0.4x_1 + 0.6x_2 + 0.9x_3 + \varepsilon \quad (87)$$

where ε has a normal distribution with mean 0 and variance 0.01. Suppose two persons A and B wish to estimate the

linear relationship for themselves. Suppose they both share the same x_0 data but they take independent measurements of variables x_1, x_2 , and x_3 . Their results are presented in a table.

A								
x_1	-3.138	-0.297	-4.582	0.301	2.729	-4.836	0.065	4.102
x_2	1.286	0.25	1.247	0.498	-0.28	0.35	0.208	1.069
x_3	0.169	0.044	0.109	0.117	0.035	-0.094	0.047	0.375
x_0	3.3979	1.6094	3.7131	1.6767	0.0419	3.3768	1.1661	0.4701
B								
x_1	-3.136	-0.296	-4.581	0.300	2.730	-4.834	0.064	4.103
x_2	1.288	0.251	1.246	0.498	-0.281	0.349	0.206	1.069
x_3	0.170	0.043	0.108	0.118	0.036	-0.093	0.048	0.376
x_0	3.3979	1.6094	3.7131	1.6767	0.0419	3.3768	1.1661	0.4701

The goal is to try and recover the actual coefficients b_0, b_1, b_2 and b_3 from the data using regression.

The reader can verify that for Person A, $\mu_0 = 1.93150$, $\sigma_{00} = 1.73426$,

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} -0.70700 \\ 0.57850 \\ 0.10025 \end{bmatrix}, \quad (88)$$

$$\mathbf{s}_0 = \begin{bmatrix} \sigma_{10} \\ \sigma_{20} \\ \sigma_{30} \end{bmatrix} = \begin{bmatrix} -3.86706 \\ 0.40169 \\ -0.048495 \end{bmatrix}, \quad (89)$$

and

$$\mathbf{S} = \begin{bmatrix} 9.33584 & -0.55747 & 0.20635 \\ -0.55747 & 0.27862 & 0.04081 \\ 0.20635 & 0.04081 & 0.01607 \end{bmatrix}. \quad (90)$$

The data suffer from multicollinearity, which is the near linear dependence of one of the variables on the remaining variables. This is evidenced by the high condition number of the covariance matrix: $\text{cond } \mathbf{S} = 3.54374 \times 10^7$. Also note that the determinant is nearly singular: $\det \mathbf{S} = 6.35047 \times 10^{-7}$.

For Person B, $\mu_0 = 1.93150$, $\sigma_{00} = 1.73426$,

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} -0.70625 \\ 0.57825 \\ 0.10075 \end{bmatrix}, \quad (91)$$

$$\mathbf{s}_0 = \begin{bmatrix} \sigma_{10} \\ \sigma_{20} \\ \sigma_{30} \end{bmatrix} = \begin{bmatrix} -3.86644 \\ 0.40204 \\ -0.04886 \end{bmatrix}, \quad (92)$$

and

$$\mathbf{S} = \begin{bmatrix} 9.33332 & -0.55747 & 0.20721 \\ -0.55747 & 0.27918 & 0.04078 \\ 0.20721 & 0.04078 & 0.01609 \end{bmatrix}. \quad (93)$$

Again the covariance matrix is ill-conditioned and nearly singular: $\text{cond } \mathbf{S} = 1.28134 \times 10^7$ and $\det \mathbf{S} = 1.75906 \times 10^{-6}$.

The standard OLS $x_0 | \{x_1, x_2, x_3\}$ regression planes for the data of Person A and Person B are

$$\text{A} : x_0 = 1.2546 + 0.9741x_1 + 9.0219x_2 - 38.4400x_3 \quad (94)$$

$$\text{B} : x_0 = 1.2752 + 0.2470x_1 + 4.5116x_2 - 17.6486x_3. \quad (95)$$

The discrepancy between the OLS regression coefficients and the model coefficients is striking. The OLS regression

coefficients disagree with the model coefficients in both sign and magnitude. Therefore OLS regression does not appear to be useful in a case such as this. In comparison, several generalized regression methods presented next appear to do a much better job in recovering the model coefficients, agreeing with the model coefficients in both sign and magnitude.

The AMR planes are

$$A : x_0 = 1.2479 - 0.4069x_1 + 0.5151x_2 + 0.9767x_3 \quad (96)$$

$$B : x_0 = 1.2478 - 0.4070x_1 + 0.5151x_2 + 0.9766x_3. \quad (97)$$

The GMR planes are

$$A : x_0 = 1.2440 - 0.4345x_1 + 0.3438x_2 + 1.8095x_3 \quad (98)$$

$$B : x_0 = 1.2433 - 0.4348x_1 + 0.3437x_2 + 1.8100x_3. \quad (99)$$

The hybrid LVR planes are

$$A : x_0 = 1.2206 - 0.4370x_1 + 0.3617x_2 + 1.9225x_3 \quad (100)$$

$$B : x_0 = 1.2200 - 0.4372x_1 + 0.3613x_2 + 1.9231x_3. \quad (101)$$

The LVR planes are computed by Tofallis as

$$A : x_0 = 1.20 - 0.43x_1 + 0.37x_2 + 1.97x_3 \quad (102)$$

$$B : x_0 = 1.20 - 0.43x_1 + 0.37x_2 + 1.98x_3. \quad (103)$$

In this example, AMR comes the closest to recovering the model coefficients. GMR, hybrid LVR and LVR appear to perform comparably well. The calculations of the three remaining OLS regressions and HMR require further study. They appear to have a negative Hessian determinant, making them inadmissible. The hybrid LVR coefficients obtained here are in good agreement with the LVR coefficients presented by Tofallis, differing only in the hundredths place. This suggests that hybrid LVR can be a useful least-squares alternative to LVR.

IV. SUMMARY

The extension of the bivariate theory of generalized least-squares to multivariate regression is begun in this paper. The multivariate symmetric least-squares problem in $m + 1$ variables seeks an m dimensional hyperplane which minimizes the average generalized mean of the square deviations between the data and hyperplane in each of the variables. The weighted multivariate ordinary least-squares problem in $m + 1$ variables is also defined using an explicit formula for the ordinary least-squares error. As in the bivariate case, every symmetric least-squares problem is shown to be equivalent to a weighted ordinary least-squares problem. The weight function $g(b_1, \dots, b_m)$ characterizes the regression method. The multivariate generalized least-squares error is then a product of the weight

function g and the explicit multivariate error function. Partial derivatives of this analytic expression for the error are then taken with respect to each of the coefficients b_1, \dots, b_m and set equal to zero. The result is a nonlinear system of equations in b_1, \dots, b_m involving only the covariances σ_{jk} which can then be solved to yield the regression coefficients for any generalized least-squares method. In order for the solution to minimize the error function and be admissible, the Hessian matrix must be computed and found to be positive definite.

The specific system of equations for the regression coefficients is presented for OLS, HMR, GMR, and AMR, for three and four variables. A related least-squares alternative to Tofallis' least-volume regression (LVR) called hybrid LVR is presented here as well.

Numerical evidence suggests that when the data are ill-conditioned ordinary least-squares regressions may not succeed in uncovering an underlying linear model. In comparison, certain generalized least-squares methods can come closer to uncovering the model coefficients.

REFERENCES

- [1] D. A. Belsley, *Conditioning Diagnostics*, New York: John Wiley & Sons, 1991.
- [2] N. R. Draper and Y. Yang, "Generalization of the mean functional relationship," *Computational Statistics and Data Analysis*, vol. 23, pp. 355-372, 1997.
- [3] S. C. Ehrenberg, "Deriving the Least-Squares Regression Equation," *The American Statistician*, vol. 37, p.232, Aug. 1983.
- [4] N. Greene, "Generalized Least-Squares Regressions I: Efficient Derivations," in *Proceedings of the 1st International Conference on Computational Science and Engineering (CSE'13)*, Valencia, Spain, 2013, pp. 145-158.
- [5] N. Greene, "Generalized Least-Squares Regressions II: Theory and Classification," in *Proceedings of the 1st International Conference on Computational Science and Engineering (CSE '13)*, Valencia, Spain, 2013, pp. 159-166.
- [6] N. Greene, "Generalized Least-Squares Regressions III: Further Theory and Classification," in *Proceedings of the 5th International Conference on Applied Mathematics and Informatics (AMATHI '14)*, Cambridge, MA, 2014, pp. 34-38.
- [7] N. Greene, "Generalized Least-Squares Regressions IV: Theory and Classification Using Generalized Means," in *Mathematics and Computers in Science and Industry*, Varna, Bulgaria, 2014, pp. 19-35.
- [8] A. Hald, *Statistical Theory with Engineering Applications*, New York: John Wiley & Sons, 1952.
- [9] C. Tofallis, "Model Fitting Using the Least Volume Criterion," in *Algorithms for Approximation IV*, J.C. Mason and J. Levesly, Eds., University of Huddersfield Press, 2002.
- [10] C. Tofallis, "Model Fitting for Multiple Variables by Minimizing the Geometric Mean Deviation," in *Total least squares and errors-in-variables modelling: algorithms analysis and applications*, S. Van Huffel and P. Lemmerling, Eds., Dordrecht: Kluwer Academic Publishers, 2002.
- [11] C. Tofallis, "Multiple Neutral Data Fitting," *Annals of Operations Research*, vol. 124, pp. 69-79, 2003.