Compound and Constrained Regression Analyses

A Dissertation Presented

by

Ling Leng

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

December 2009

Stony Brook University

The Graduate School

Ling Leng

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Wei Zhu—Dissertation Advisor

Professor, Department of Applied Mathematics and Statistics

Nancy Mendell—Chairperson of Defense

Professor, Department of Applied Mathematics and Statistics

Stephen Finch

Professor, Department of Applied Mathematics and Statistics

Ellen Li

Professor, Department of Medicine, Stony Brook University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Compound and Constrained Regression Analyses

by

Ling Leng

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2009

In linear regression analysis, randomness often exists in both the dependent and the independent variables and the resulting model is referred to as the error in variable (EIV) model. Estimation of the regression parameters using the current EIV structural model approach is dependent upon the ratio of the error variances, which is usually unknown. Furthermore, the current structural model approach is a parametric approach assuming normal distributions for all random variables involved. To overcome these impasses, we introduce two alternative frameworks, the compound regression analysis and the constrained regression analysis methods. It is shown that these approaches are equivalent to each other and, to the parametric structural model approach when the random variables involved are normally distributed. The advantages of the new regression approaches lie in their intuitive geometric representations, their distribution free non-parametric nature being a direct generalization of the ordinary least squares method, and their operational independence to the ratio of the error variances. Examples and simulations are provided to motivate and to illustrate these new approaches.

Table of Contents

List of Figures

List of Tables

1. Introduction

The classical ordinary least squares regression theory relies on the assumption that the explanatory variables are measured without error which is often untrue in climate modelling and other scientific research. In gauging the relationship between the concentrations of organic aerosols and anthropogenic carbon monoxide (CO) (Jobson et al. 1999; and Kleinman et al. 2007), we found that both quantities, measured by the mass spectrometer and the UV fluorescence analyzer respectively, contain measurement errors and possibly other volatilities due to air dynamics as well. Two commonly used regression methods for simple linear regression with a random regressor, the orthogonal regression and the geometric mean regression, yielded different regression equations, as expected, for our dataset. The immediate question confronting the scientist is which one to adopt for his/her study. This is a typical example of the error in variable (EIV) regression, also known as the measurement error model, where both the dependent and the independent variables contain unknown errors. In the Frequentist context, this is still an outstanding problem. As E. T. Jaynes pointed out in his celebrated monograph *Probability Theory – The logic of science*:

“As science progressed to more and more complicated problems of inference, the shortcoming of the orthodox methods became more and more troublesome. Fisher would have been nearly helpless, and Neyman completely helpless, in a problem with many nuisance parameters but no sufficient or ancillary statistics. Accordingly, neither ever attempted to deal with what is actually the most common problem of inference faced by experimental scientists: linear regression with both variables subject to unknown error." (Jaynes 2004; page 497)

To date, the general frequentist approach to EIV modeling is through the maximum likelihood estimation method (Lindley 1947; Wong 1989). However, this method depends on the multivariate normality assumptions and furthermore, to make the matter worse, the solution depends on the ratio of the error variances, which is usually unknown.

To overcome these impasses for EIV regression, I have developed two novel general regression approaches -- the compound and the constrained regression analysis methods that will provide intuitive and practical solutions for all EIV regression problems including our own. My approach is non-parametric and includes both the orthogonal regression and the geometric mean regression as special cases. When the multivariate normality assumption holds, our methods will produce identical solutions as the traditional MLE method. However, an added advantage of our approaches is that we can circumvent the unknown variance ratio problem and yield optimal or near optimal solutions in the absence of such knowledge. This thesis is organized as follows. The new methods are introduced in Section 3 and illustrated, through three examples and one simulation study, in Section 4. Discussion and future work directions are presented in Section 5. In the following, we begin by a brief review of the current approaches.

2. EXISTING METHOD

A general structural model approach for simple linear regression when both variables are random, that is, the error in variable (EIV) model, is as follows (Sprent 1969; Wong 1989).

$$\begin{aligned} X &= \xi + \delta \qquad \delta \sim N\left(0, \sigma_\delta^2\right) \\ Y &= \eta + \varepsilon \qquad \varepsilon \sim N\left(0, \sigma_\varepsilon^2\right) \\ \eta &= \beta_0 + \beta_1 \xi \end{aligned}$$

Here δ and ε are independent random errors. There are two analysis approaches concerning this model: the functional and the structural. Their basic difference is whether to consider ξ as a non-random variable or a random variable following normal distribution with mean μ and variance τ^2, and independent to both random errors. Since the latter approach is more general, in the discussion below, we will follow the structural model approach where X and Y follow a bivariate normal distribution:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ \beta_0 + \beta_1\mu \end{pmatrix}, \begin{pmatrix} \tau^2 + \sigma_\delta^2 & \beta_1\tau^2 \\ \beta_1\tau^2 & \beta_1^2\tau^2 + \sigma_\varepsilon^2 \end{pmatrix}\right)$$

Given a random sample of observed X's and Y's, the maximum likelihood estimator (MLE) of the regression slope is given by

$$\hat{\beta}_1 = \frac{S_{YY} - \lambda S_{XX} + \sqrt{(S_{YY} - \lambda S_{XX})^2 + 4\lambda S_{XY}^2}}{2S_{XY}}$$

Its value depends on the ratio of the two error variances $\lambda = \sigma_\varepsilon^2 / \sigma_\delta^2$, which is generally unknown and unable to be estimated from the data alone (Lindley 1947).

It has been shown that the ordinary least squares regressions (OLS) and the two most commonly used regression methods when both X and Y are random, the orthogonal regression (OR) and the geometric mean regression (GMR), can be considered special cases in this structural model approach, with the distinction that these specific methods do not rely on the bivariate normal assumption.

The OLS slope estimator with Y or X as the dependent variable will minimize the squared vertical or horizontal distances from the points to the regression line, and corresponds to the MLE of the slope in the structural model approach when $\lambda = \infty$ or $\lambda = 0$. The OLS is suitable when only one of the two variables is random.

When both variables are random, the orthogonal regression takes the middle ground by minimizing the sum of squared orthogonal distances from the observed data points to the regression line. The resulting estimator of the slope is (Jackson and Dunlevy 1988):

$$\hat{\beta}_1 = \frac{S_{YY} - S_{XX} + \sqrt{(S_{YY} - S_{XX})^2 + 4S_{XY}{}^2}}{2S_{XY}}$$

It is the same as the MLE in the structural model approach when $\lambda = 1$. This means that the OR is suitable when the error variances are equal.

The geometric mean regression, on the other hand, took the middle ground by taking the geometric mean of the slope of y on x, and the reciprocal of the slope of x on y OLS regression lines resulting in the estimated slope

$$\hat{\beta}_1 = sign(S_{XY})\sqrt{\hat{\beta}_{OLS,\text{ Y on X}} * (\hat{\beta}_{OLS,\text{ X on Y}})^{-1}} = sign(S_{XY})\sqrt{\frac{S_{YY}}{S_{XX}}}$$

The GMR can also be obtained by minimizing the sum of the triangular areas bounded by the vertical and the horizontal projections from the data points to the regression line and the regression line itself (Barker et al. 1988).

Comparing to the structural model approach, the GMR estimator is equal to the MLE when $\lambda = S_{YY}/S_{XX}$ (Sprent and Dolby 1980). This means that the GMR approach is suitable when the randomness from X and Y are from the random errors only. That is, when we take the functional analysis approach by assuming that ξ is not random.

When the ratio of error variances is known, the MLE of the slope parameter can be extended to multivariate case. Consider the multivariate linear relationship

$$\sum_{j=1}^{p} \beta_j \xi_j = \alpha \qquad (1)$$

between p variables ξ_1, ξ_2,..., ξ_p where observations $x_1, x_2, ..., x_p$ are made with independent normally distributed errors. Hence

$$x_i = \xi_j + \varepsilon_j$$

while ε_j are independent of each other and follow a normal distribution with variance equal to λ_p and $(\xi_1, \xi_2, ..., \xi_p)$ follows a p-variate normal distribution $N(\mu_s, \Sigma_s)$. Therefore, the observations $x = \xi + \varepsilon$ have the distribution $N_p(\mu_s, \Sigma_s + \sigma^2\Lambda)$.

The maximum likelihood estimate of parameter can be obtained by

$$\hat{\beta} = (\hat{\varpi}'_1 \Lambda^{-1} \hat{\varpi}_1)^{-\frac{1}{2}} \Lambda^{-\frac{1}{2}} \hat{\varpi}_1 \qquad (2)$$

where $\hat{\varpi}_1$ is the eigenvector corresponding to the smallest eigenvalue of $\Lambda^{-\frac{1}{2}} S \Lambda^{-\frac{1}{2}}$ and S is the sample variance-covariance matrix of X (W.M. Patefield 1981).

Orthogonal regression also has an extension to the higher dimension case. The distance from a point (x_0, y_0, z_0) to a regression plane $\beta_1 X + \beta_2 Y + \beta_3 Z = c$ is:

$$\frac{(\beta_1 x_0 + \beta_2 y_0 + \beta_3 z_0 - c)^2}{\beta_1^2 + \beta_2^2 + \beta_3^2}$$

Hence, the orthogonal regression will minimize the following formula:

$$\sum_{i=1}^{n} \frac{(\beta_1 x_i + \beta_2 y_i + \beta_3 z_i - c)^2}{\beta_1^2 + \beta_2^2 + \beta_3^2}$$

Then there is a formula for k-dimensions:

$$\sum_{i=1}^{n} \frac{(\beta_1 X_{1i} + \beta_2 X_{2i} + \ldots \beta_k X_{ki} - c)^2}{\beta_1^2 + \beta_2^2 + \ldots + \beta_k^2}$$

We can prove that orthogonal regression is a special case of GLS when all λ s are equal to 1(see Appendix A theorem1).

3. COMPOUND AND CONSTRAINED REGRESSION ANALYSES

The structural model approach has two fundamental difficulties for real life applications. First, it requires the variables to follow a joint bivariate normal distribution. Second, it require knowledge of λ --the ratio of the error variances, which is usually unknown and cannot be estimated from the data statistically (Lindley 1947; Wong 1989). In addition to these impediments, the structural model approach has also lost the intuitive geometric interpretations enjoyed by the other, albeit more specialized methods such as

OLS, OR or GMR.

In this section, we present the novel compound regression analysis and the constrained regression analysis methods. Both approaches enjoy clear geometric interpretations. Furthermore, we prove that they are equivalent to each other, and to the structural model approach when the joint distribution is bivariate normal. The added benefits are that, firstly, the estimators from these new approaches can be derived using the ordinary least squares method without the normality assumption. Secondly, users can choose their desirable compound/constrained regression line without the knowledge of λ, the usually unknown and un-estimable ratio of the error variances.

3.1 Compound Regression Analysis

The idea of compound regression analysis came from our experience with multiple-objective optimal designs (Biedermann et al. 2006; Dette et al. 2005; Zhu 1996; Zhu and Wong 1998). When we have only one objective in an experiment, the optimal design ξ, defined as a probability mass function that places total mass on a finite collection of *k* points in the design region X, is derived by minimizing a convex function of the Fisher Information matrix corresponding to this objective. When we have two objectives in mind represented by the convex functions Φ_1 and Φ_2 respectively, a standard and intuitive approach is to find a compound optimal design (L äuter 1974, 1976) that will minimize a linear combination of these objective functions $\Phi_\gamma = \gamma\Phi_1 + (1-\gamma)\Phi_2$,

$0 \le \gamma \le 1$. The value of γ is determined using the concept of design efficiency, $e_i = \Phi_i^* / \Phi_i$ - where Φ_i^* is the optimum (minimum) value for the i^{th} objective alone $(i = 1, 2)$. The design efficiency gauges how efficient the given design is for estimating each objective. A user can plot both efficiencies for all possible values of γ, $0 \le \gamma \le 1$, to decide the γ value corresponding to the desirable efficiency values. Such a plot is called an efficiency plot.

This idea of molding two objectives in a design setting through a compound criterion can be readily applied to the scenario of simple linear regression when both variables are random. For the OLS on Y and X separately, variation exists in the Y or X direction only and thus one would minimize the sum of squared distances along the vertical or horizontal axis only to obtain the best regression line for each scenario. When both Y and X are random, one would naturally wish to find a regression line $Y = \beta_0 + \beta_1 X$ that will minimize variations in both directions.

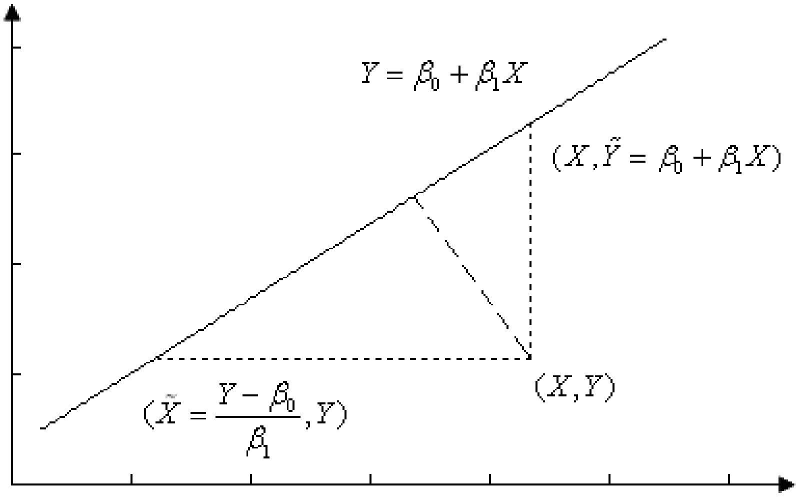

$Y = \beta_0 + \beta_1 X$
$(X, \tilde{Y} = \beta_0 + \beta_1 X)$
(X, Y)
$(\tilde{X} = \frac{Y - \beta_0}{\beta_1}, Y)$

Figure 1. Illustration of the Ordinary Least Squares Regressions, Orthogonal Regression, Geometric Mean Regression, Compound Regression and Constraint Regression Analyses.

This can be accomplished by minimizing a weighted average of the squared vertical and horizontal distances, as illustrated in Figure 1, as follows:

$$\begin{aligned} SS_{\gamma} &= \gamma \sum_{i=1}^{n} (Y_i - \tilde{Y}_i)^2 + (1-\gamma) \sum_{i=1}^{n} (X_i - \tilde{X}_i)^2 \\ &= \gamma \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 + (1-\gamma) \sum_{i=1}^{n} (X_i - \frac{Y_i - \beta_0}{\beta_1})^2, \qquad 0 \le \gamma \le 1. \end{aligned}$$

At the two extreme values of $\gamma = 1$ and $\gamma = 0$, we obtain the OLS on Y or X respectively. For each γ, we can obtain the least squares estimators of the regression parameters by solving $\frac{\partial SS_r}{\partial \beta_0} = 0$ and $\frac{\partial SS_r}{\partial \beta_1} = 0$ simultaneously. Straight-forward derivation shows that the resulting compound regression model estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ would satisfy

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} \quad \text{and} \quad \frac{\gamma}{1-\gamma} \beta_1^4 S_{XX} - \frac{\gamma}{1-\gamma} \beta_1^3 S_{XY} + \beta_1 S_{XY} - S_{YY} = 0 \qquad (3)$$

Solutions can be obtained using any standard numerical software such as MATLAB.

For the higher dimension case, the process can be carried out in the same way. The compound regression takes account of all the prediction errors with different weight and gets the slope estimate by minimizing the following sum function

$$SS_{\gamma} = \gamma_1 \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \gamma_2 \sum_{i=1}^{n} (x_{1i} - \hat{x}_{1i})^2 + \cdots + \gamma_k \sum_{i=1}^{n} (x_{ki} - \hat{x}_{ki})^2$$

In order to do this, we simplify the above sum function

$$SS_\gamma = \gamma_1 \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 + \gamma_2 \frac{1}{\beta_1^2} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 + \gamma_k \frac{1}{\beta_k^2} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

$$= (\gamma_1 + \frac{\gamma_2}{\beta_1^2} + \cdots + \frac{\gamma_k}{\beta_k^2}) \sum_{i=1}^{n} (y_i - \overline{y} - \beta_1 (x_1 - \overline{x}_1) - \cdots \beta_2 (x_k - \overline{x}_k))^2$$

$$= (\gamma_1 + \frac{\gamma_2}{\beta_1^2} + \cdots + \frac{\gamma_k}{\beta_k^2})(SYY + \sum_{i=1}^{k} \sum_{j=1}^{k} \beta_i \beta_j SX_i X_j - 2 \sum_{i=1}^{k} \beta_i SX_i Y)$$

And set derivatives:

$$\frac{\partial SS_\gamma}{\partial \beta_i} = (\gamma_1 + \frac{\gamma_2}{\beta_1^2} + \cdots + \frac{\gamma_k}{\beta_k^2})(\sum_{\substack{j=1 \\ j \neq i}}^{k} \beta_j SX_i X_j - 2 \sum_{i=1}^{k} SX_i Y + 2\beta_i SX_i X_i)$$
$$- \frac{2\gamma_i}{\beta_i^3} (SYY + \sum_{i=1}^{k} \sum_{j=1}^{k} \beta_i \beta_j SX_i X_j - 2 \sum_{i=1}^{k} \beta_i SX_i Y) = 0 \qquad (4)$$

to obtain the slope estimates.

3.2 Equivalence between Compound Regression and Structural Model

In this section, we prove that there is a one-to-one correspondence between the MLE in the structural model approach (under different λ) and the least squares estimator in compound regression analysis (under different γ) for the slope parameter β, and thus for the corresponding regression line/plane.

Theorem 2. The compound regression analysis and the structural model approach are equivalent to each other.

Proof:

a) From formula (2) above, we can obtain $\hat{\beta}$ from variances ratio λ, then we can plug it into equations (4) to get the corresponding γ_1 and γ_2.

b) We can get the slope estimator $\hat{\beta}$ from compound regression. We define Λ using the formula $\Lambda^{-1}S\hat{\beta}=\lambda\hat{\beta}$, and define $\hat{\omega}_1=\Lambda^{-\frac{1}{2}}\hat{\beta}$. We will have the following equivalence.

$$\Lambda^{-\frac{1}{2}}S\Lambda^{-\frac{1}{2}}\hat{\omega}_1=\lambda\hat{\omega}_1 \leftrightarrow \Lambda^{-1}S\Lambda^{-\frac{1}{2}}\hat{\omega}_1=\lambda\Lambda^{-\frac{1}{2}}\hat{\omega}_1 \leftrightarrow \Lambda^{-1}S\hat{\beta}=\lambda\hat{\beta}$$

We can see this is the same with MLE of GLS (actually λ does not matter here, what we concern is only the ratio of λ s to each other, not the actual number). Hence, the Λ we define here can yield the same slope estimate of compound regression. Therefore, we have proven the existence of Λ. That is, we have proven the equivalence. □

Now that we have shown the equivalence between the structural model and the compound regression approaches, our problem transfers from finding the desirable regression line/space from a class of unknown $\lambda's$ to a class of unknown $\gamma's$. The constrained regression analysis method will further elucidate the path to the solution.

3.3. Constrained Regression Analysis

In experimental design, besides the compound optimal design approach, the other way to derive the best design satisfying multiple research objectives is through the constrained optimal design approach (Lee 1987, 1988). Intrigued by the similarity between the multiple-objective design and the random regressor scenarios, we have devised the compound regression analysis approach as shown in the previous sections. Here we derive the constrained regression analysis, the regression counterpart of the constrained optimal design. And, just as the two approaches are shown to be equivalent to each other in the optimal design scenario (Cook and Wong 1994; and Clyde and Chaloner 1996), we find them equivalent in the regression scenario as well.

We define the constrained regression as follows. Given the constraint of $\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2 \le c$ where c is a user selected non-negative constant, the compound regression line will minimize $\sum_{i=1}^{n}(X_i-\tilde{X}_i)^2$.

Theorem 3. The constrained regression is equivalent to the compound regression in that there is a 1-1 correspondence between $c\left(c \ge 0\right)$ and γ $\left(0 \le \gamma \le 1\right)$. For a given c we have $\gamma = \dfrac{S_{YY}-\hat{\beta}_1 S_{XY}}{S_{YY}-\hat{\beta}_1 S_{XY}+\hat{\beta}_1^4 S_{XX}-\hat{\beta}_1^3 S_{XY}}$ and $\hat{\beta}_1 = \dfrac{S_{XY}+sign(S_{XY})\sqrt{S_{XY}^2-S_{XX}(S_{YY}-c)}}{S_{XX}}$.

Proof of this theorem is provided in the Appendix A.

The advantage of the constrained regression analysis approach lies in its intuitive interpretation. The constrained regression can be stated equivalently in terms of the regression efficiencies for Y and X defined as

$$e_1 = \frac{\min\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2}{\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2} = \frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i^{OLS(Y)})^2}{\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2} \quad \text{and} \quad e_2 = \frac{\min\sum_{i=1}^{n}(X_i-\tilde{X}_i)^2}{\sum_{i=1}^{n}(X_i-\tilde{X}_i)^2} = \frac{\sum_{i=1}^{n}(X_i-\hat{X}_i^{OLS(X)})^2}{\sum_{i=1}^{n}(X_i-\tilde{X})^2}$$

respectively. For a given $c^* \in \left[0,1\right]$, the constrained regression line will maximize e_2 subject to $e_1 \ge c^*$.

With the equivalence of the constrained and the compound regression approaches, we can first calculate all the compound regression lines given that they are computationally more efficient than their constrained regression counterparts. Then we plot the efficiency curves for all possible $\gamma\ \left(0 \leq \gamma \leq 1\right)$, and select, from which, the value of γ^* corresponding to a desired c^* (and thus a desirable constrained regression line with intuitive interpretations). The intersection of the line $\gamma = \gamma^*$ and the curve of e_2 in the efficiency plot would yield the best efficiency we can achieve for the estimation of X.

Here we point out that both new regression approaches are symmetric for the estimations of X and Y and thus, one can reverse the order of the importance for X and Y and obtain the best regression line for Y subject to $e_2 \geq c^{**}$. Now that we have successfully circumvented the dilemma of the unknown error variance ratio λ, we will demonstrate our approaches with examples in the following section.

For higher dimensional case, constrained regression will follow the same form. We can give the constraint on $\sum_{i=1}^{n}(Y_i - \tilde{Y}_i)^2 \leq c_0, \sum_{i=1}^{n}(X_{ji} - \tilde{X}_{ji})^2 \leq c_j, j = 1,..,k-1,k+1,..,p$ and calculate the slope estimates that will minimize $\sum_{i=1}^{n}(X_{ki} - \tilde{X}_{ki})^2$. Or alternatively, we can define the efficiency of regression:

$$e_0 = \frac{\min \sum_{i=1}^{n}(Y_i - \tilde{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \tilde{Y}_i)^2} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i^{OLS(Y)})^2}{\sum_{i=1}^{n}(Y_i - \tilde{Y}_i)^2}$$

$$e_j = \frac{\min \sum_{i=1}^{n} (X_{ji} - \tilde{X}_{ji})^2}{\sum_{i=1}^{n} (X_{ji} - \tilde{X}_{ji})^2} = \frac{\sum_{i=1}^{n} (X_{ji} - \hat{X}_{ji}^{OLS(X_j)})^2}{\sum_{i=1}^{n} (X_{ji} - \tilde{X}_{ji})^2}, j = 1,..,k-1,k+1,..,p$$

And under the constraints $e_0 \geq c_0, e_j \geq c_j, j = 1,..,k-1,k+1,..,p$, we calculate the slope estimates that will maximize the efficiency of regression of x_k.

Although we have proven the equivalence in two-dimensional case, the theoretical proof of extension to higher dimension seems difficult to obtain. However, we did some simulation to find the numerical solution of constrained regression in examples.

4. EXAMPLES

Our first example is classic; our second example is from the atmospheric science that has motivated this work and the third example is from a gene microarray study that serves to illustrate EIV model in higher dimension. That is, when we deal with multiple regressions with more than one random regressors.

4.1 Example 1

Error in both variables problem has been noticed for a long time. Casella and Berger (2001, pages 542, 579) introduced this classic example where randomness exists in both directions. Figure 2 shows the scatter plot of the data together with some typical regression lines. From the previous discussion, we know that the entire class of compound regression lines would range from OLS(X) to OLS(Y). The question is which regression line to choose among this diversified class of regression models.

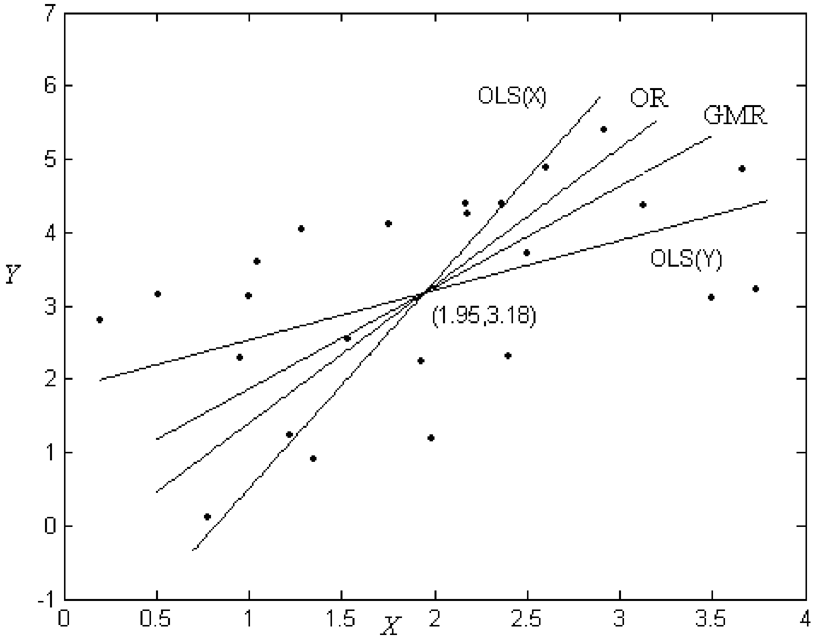


Figure 2. Span of Regression Lines for Example 1.

Table 1. Selected Compound Regression Analysis Results for Example 1.

γ	$\hat{\beta}_1$	$\hat{\beta}_0$	$\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2$	$\sum_{i=1}^{n}(X_i-\tilde{X}_i)^2$	e_1	e_2	λ
0 (OLS_X)	2.82	−2.31	137.53	17.33	0.24	1.00	0.00
0.07(OR)	1.88	−0.48	65.87	18.71	0.50	0.93	1.00
0.10	1.79	−0.30	61.09	19.16	0.54	0.90	1.13
0.20	1.57	0.13	51.06	20.84	0.65	0.83	1.50
0.30	1.43	0.39	46.00	22.50	0.72	0.77	1.79
0.34(GMR)	1.39	0.48	44.43	23.24	0.75	0.75	1.91
0.40	1.33	0.59	42.72	24.25	0.78	0.71	2.07
0.50	1.24	0.76	40.31	26.22	0.82	0.66	2.36
0.60	1.16	0.92	38.40	28.56	0.86	0.61	2.71

0.70	1.08	1.08	36.79	31.56	0.90	0.55	3.17
0.80	0.99	1.24	35.39	35.84	0.94	0.48	3.90
0.90	0.89	1.45	34.11	43.35	0.97	0.40	5.57
1（OLS_Y）	0.68	1.86	33.12	71.94	1.00	0.24	∞

Table 1 above tabulates selected compound regression lines including OLS(X), OLS(Y), OR and GMR. The efficiencies for estimating X and Y ranging from 0.24 to 1 in opposite directions as γ goes from 0 to 1. The OR line is more efficient in reducing variations in the X direction than the Y direction with efficiencies for X and Y being 0.93 and 0.50 respectively. The GMR, however, provides a nice balance between the two estimations yielding equal efficiencies (0.75) for both X and Y. Such is not a mere coincidence; in fact, it is universally true as stated in the following theorem with proof provided in the Appendix.

Theorem 4. (a) The Geometric Mean Regression would always yield equal efficiencies for the estimations of X and Y respectively. (b) The Ordinary Least Squares Regressions for X and Y have the same efficiencies, albeit in reverse order, for X and Y.

For each given data set, users can select the desired regression line from the entire class of all compound regression lines using the efficiency plot as shown in Figure 3. The estimated regression line can be easily computed using Equation (1) in Section 3.1. Our estimations of γ and e_2 are highly accurate because of the explicit analytical formula in Theorem 2. Suppose that the user want the desired line to be at least 95% efficient for the estimation of Y. From this plot, it is easy to see that when $e_1 = 0.95$, we have

$\gamma = 0.8401$ and $e_2 = 0.453$. If, however the user wishes the desired line to be at least 85% efficient for the estimation of Y. From the efficiency plot, we will find that $e_1 = 0.85$ corresponds to $\gamma = 0.5686$ and $e_2 = 0.624$.

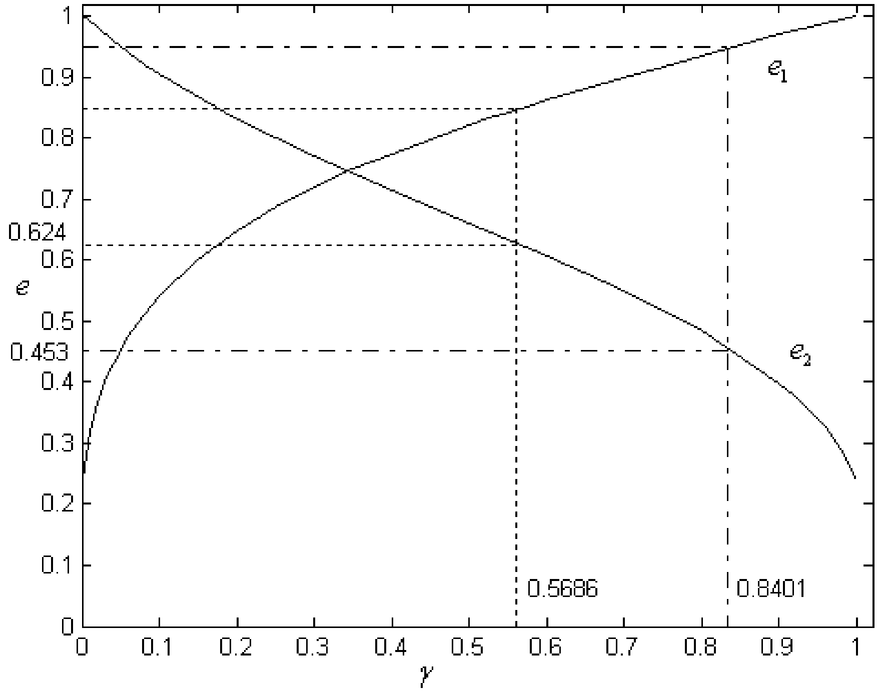


Figure 3. Efficiency plot for Example 1.

4.2 Example 2

Our second example from the atmospheric science also motivated this work. In order to investigate the organic aerosol evolution, the time evolution of aerosol concentration and chemical composition in a megacity urban plume was determined based on 8 flights of the DOE G-1 aircraft in and downwind of Mexico City during the March 2006 MILAGRO field campaign (Kleinman et al.2007). The data consist of 113 pairs of carbon monoxide (CO) and organic aerosol concentrations observed above the

Mexico City. Our goal is to study the linear relationship between the random variables ln(CO), as X, and the concentration of organic aerosol (Y). Similar to Example 1, we can visually inspect the span of compound regression lines ranging from OLS(X) to OLS(Y) in the following scatter plot.

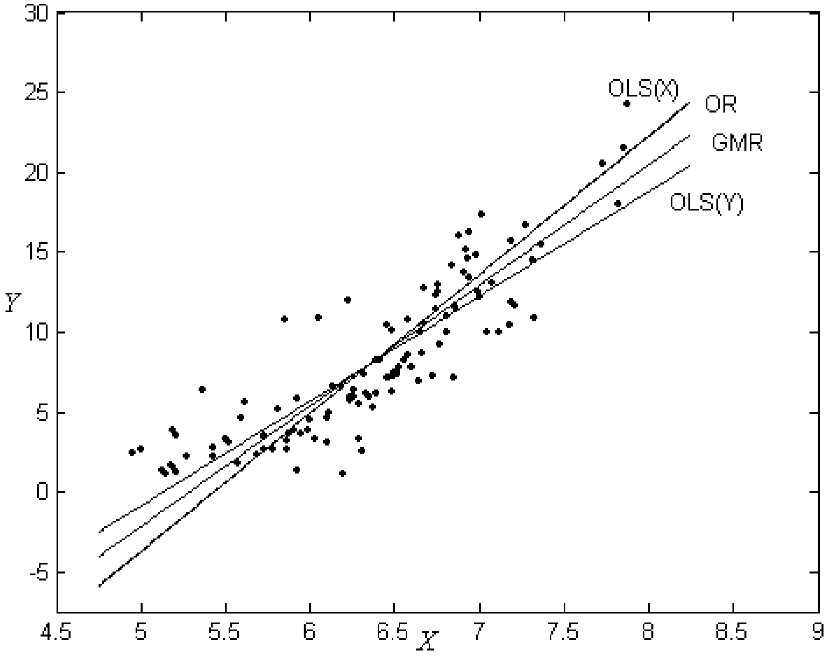


Figure 4. Span of Regression Lines for Example 2.

Unlike Example 1 where the scales of the two random variables are comparable, here the scale of Y can be three times as large as that of X. Subsequently the sum of squares for Y would be much larger than that for X which means the former would dominate the minimization of the compound regression sum of squares SS_γ for most γ. We will still obtain the entire class of compound regression lines however the efficiency plot would be flat on most of the interval for γ and then change abruptly at the end of the interval, which will hamper the visual inspection and selection of desired regression lines. This phenomenon can be easily corrected by standardizing the compound regression sum of squares as follows:

$$SS_{\gamma}^{sd}=\gamma\frac{\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2}{\min\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2}+(1-\gamma)\frac{\sum_{i=1}^{n}(X_i-\tilde{X}_i)^2}{\min\sum_{i=1}^{n}(X_i-\tilde{X}_i)^2}$$

$$=\gamma\frac{\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2}{\sum_{i=1}^{n}(Y_i-\hat{Y}_i^{OLS(Y)})^2}+(1-\gamma)\frac{\sum_{i=1}^{n}(X_i-\tilde{X}_i)^2}{\sum_{i=1}^{n}(X_i-\hat{X}_i^{OLS(X)})^2}\qquad 0\le\gamma\le 1$$

The resulting standardized compound regression is easily shown to be equivalent to the constrained regression as well as the structural model. Selected compound regression lines using the standardized criterion are shown in Table 2.

Table 2. Selected Compound Regression Analysis Results for Example 2.

γ	β_1	β_0	$\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2$	$\sum_{i=1}^{n}(X_i-\tilde{X}_i)^2$	e_1	e_2	λ
0 （OLS_X）	8.68	−47.13	895.63	11.891	0.755	1.000	0.00
0.01(OR)	8.64	−46.90	888.08	11.892	0.761	1.000	1.00
0.10	8.36	−45.13	835.29	11.943	0.810	0.996	9.56
0.20	8.12	−43.56	795.04	12.066	0.851	0.985	19.08
0.30	7.91	−42.23	765.36	12.239	0.884	0.972	29.47
0.40	7.72	−41.03	742.24	12.458	0.911	0.955	41.62
0.50(GMR)	7.54	−39.90	723.63	12.725	0.934	0.934	56.87
0.60	7.37	−38.80	708.42	13.052	0.955	0.911	77.70
0.70	7.19	−37.68	696.01	13.459	0.972	0.884	109.73
0.80	7.01	−36.50	686.17	13.981	0.985	0.851	169.44
0.90	6.80	−35.19	679.17	14.689	0.996	0.810	338.29
1 （OLS_Y）	6.55	−33.62	676.20	15.750	1.000	0.755	∞

From Table 2, we observe that the efficiency of predicting Y increases from 0.755

to 1 while the efficiency of predicting X decreases from 1 to 0.755 as γ goes from 0 to 1. We also observe that the GMR yields equal efficiencies (0.934) for the estimations of X and Y as proven in Theorem 3. We would highly recommend this regression line to our collaborators for the given study. In case they wish for a slightly higher efficiency for the estimation of Y, say 95%, one can easily find the corresponding compound regression line with $\gamma = 0.575$ and a 92% efficiency for X as illustrated in Figure 5.

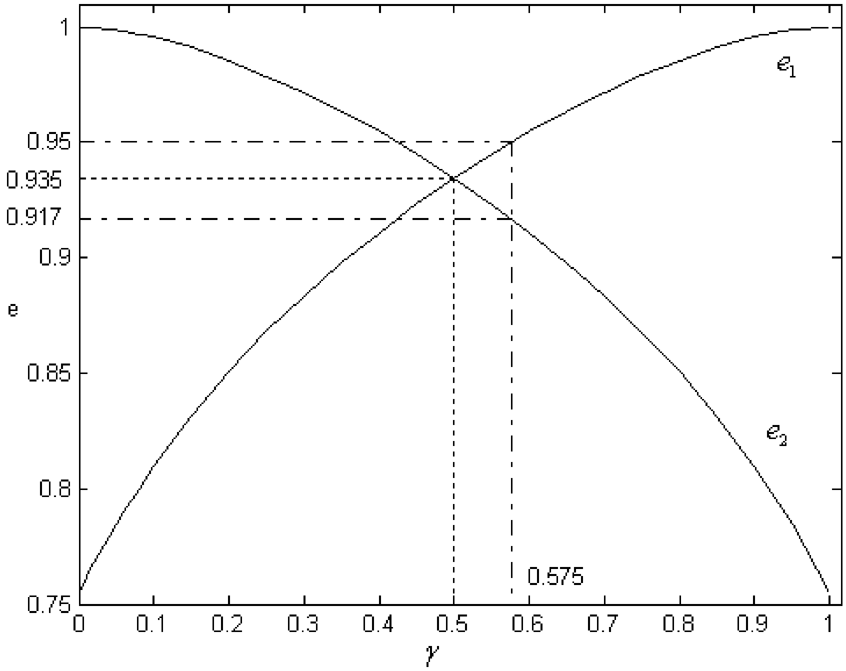


Figure 5. Efficiency plot for Example 2.

4.3 Example 3

In microarray analysis, measurement error exists in every gene expression measure. Therefore, ordinary least square regression is not suitable here. In this example, we analyze a microarray data with 95 observations. Our goal is to find the linear relationship between 3 genes for possible genetic pathway relations. Here, GLS and compound regression method are used.

First, we find that these genes are not normally distributed. Therefore, log-transformations are used to modify the data. QQ-plots are shown below:

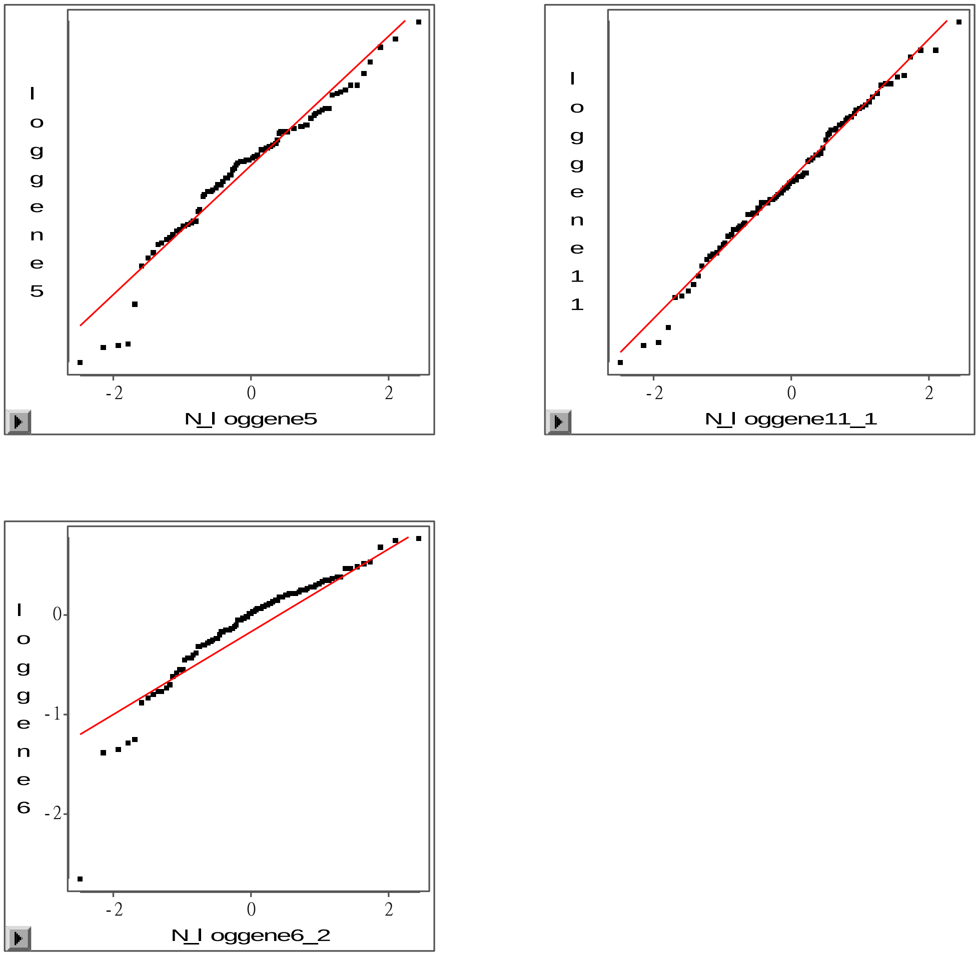


Figure 6. QQ-plots for Example3

Second, the general least square method gives up the following model for this example:

$$\log gene5 = \alpha + \beta_1 \log gene6 + \beta_2 \log gene11$$

If we have the knowledge of λ s (actually, we only need to know the ratio of λ s), we can plug this into formula (3) and get the slope estimates. For example, if the measurement errors are equal, then $\Lambda = I_3$ and the slope estimates are:

$\beta_1 = -0.2,\ \beta_2 = 1.12$ and the corresponding γ s are: $\gamma_1 = 0.3899,\ \gamma_2 = 0.0006,\ \gamma_3 = 0.6095$

However, there are cases where we don't have knowledge of the measurement errors. And since we cannot get the measurement error from the data, λ s cannot be obtained from statistics. In this case, we can use our compound regression. For different γ s, we get the following estimate:

γ=[1,0,0] the slope estimates are $\beta_1 = 0.2174,\ \beta_2 = 0.4647$. This is the same with the ordinary least square regression using model above;

γ=[0.5,0.5,0] the slope estimates are $\beta_1 = 0.8430,\ \beta_2 = -0.0985$. The corresponding λ s are $\lambda_1 = 1, \lambda_2 = 1.9804, \lambda_3 = 0.0007$;

γ=[0.33,0.33,0.33] the slope estimates are $\beta_1 = 0.5079$, $\beta_2 = 0.5809$.

Here we also give the following table stating some value of γ s and the corresponding slope estimate, λ s and efficiencies.

Table 3. Selected Compound Regression Analysis Results for Example 3.

γ_1	γ_2	γ_3	β_1	β_2	λ_1	λ_2	λ_3	e_1	e_2	e_3
1	0	0	0.2174	0.4647	1	0	0	1.0000	0.0847	0.2167
0	1	0	2.5657	-1.6494	0	1	0	0.0848	1.0000	0.2314

0	0	1	-0.7720	2.1443	0	0	1	0.2167	0.2314	1.0000
0.5	0.5	0	0.8430	-0.0985	1	1.98	0	0.5662	0.7211	0.0055
0.5	0	0.5	-0.0972	0.9988	1	0	1	0.7323	0.0124	0.7332
0	0.5	0.5	0.5604	0.6299	0	1.6	1	0.5426	0.3054	0.2160
0.33	0.33	0.33	0.5079	0.5809	1	15	9	0.6434	0.2974	0.2179
0.3899	0.0006	0.6095	-0.2000	1.1181	1	1	1(OR)	0.6452	0.0457	0.8083

For different γ s, we get different efficiencies and also get the projection plot of efficiency e_1, e_2 and e_3 on plane γ_1 v.s. γ_2

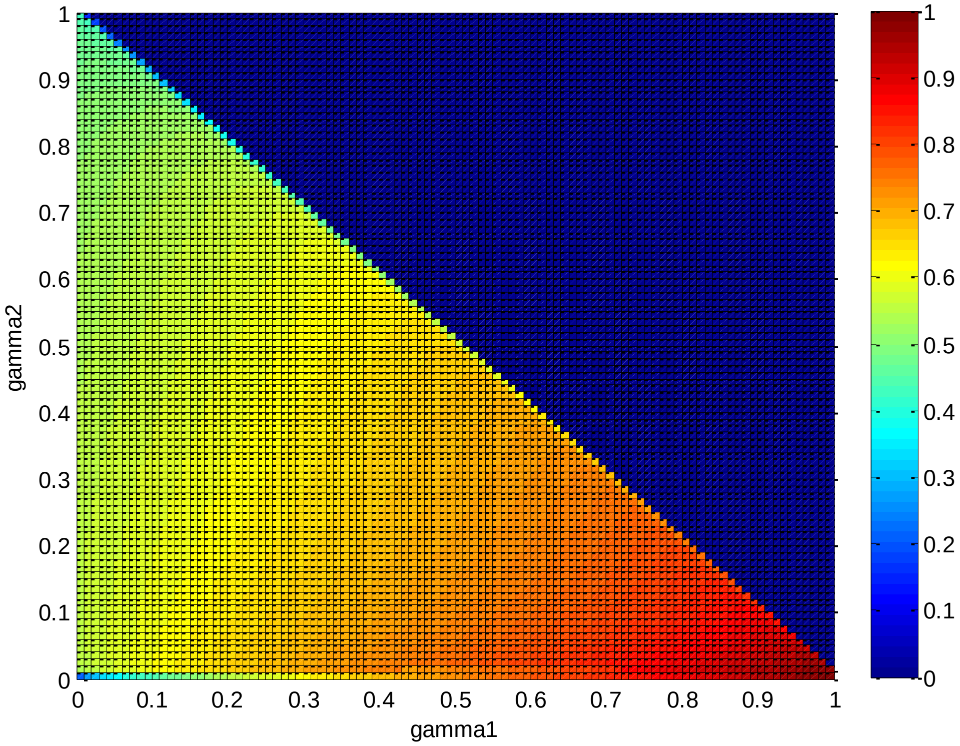


Figure7a. Efficiency plot of loggene5 (projection to γ_1-γ_2 plane)

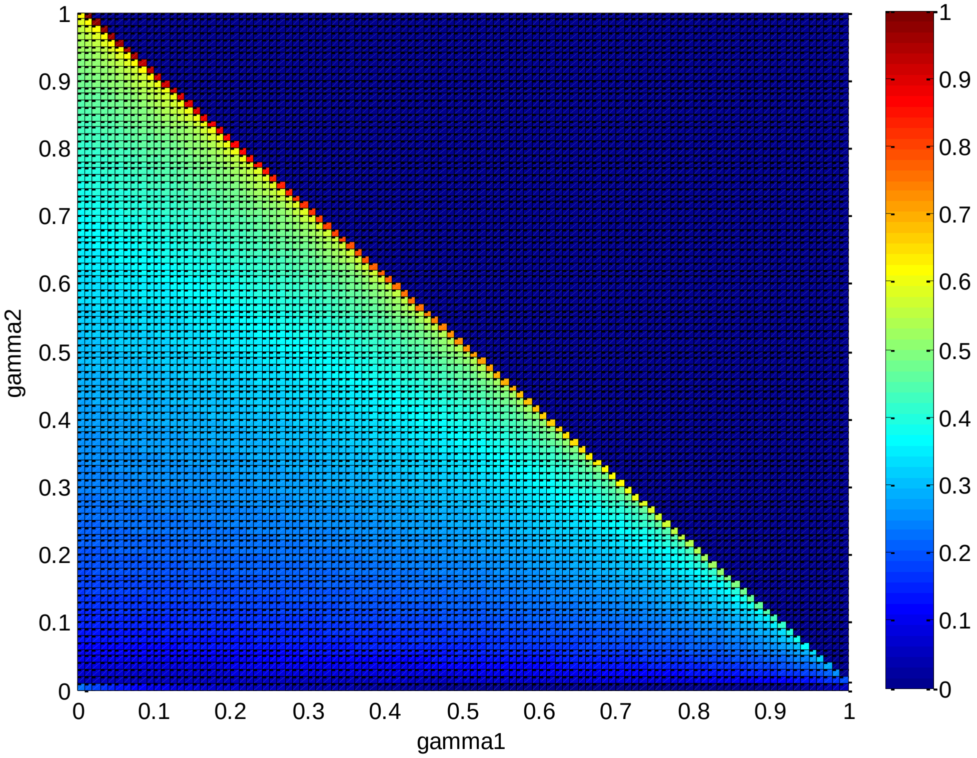


Figure 7b. Efficiency plot of loggene6 (projection to γ_1-γ_2 plane)

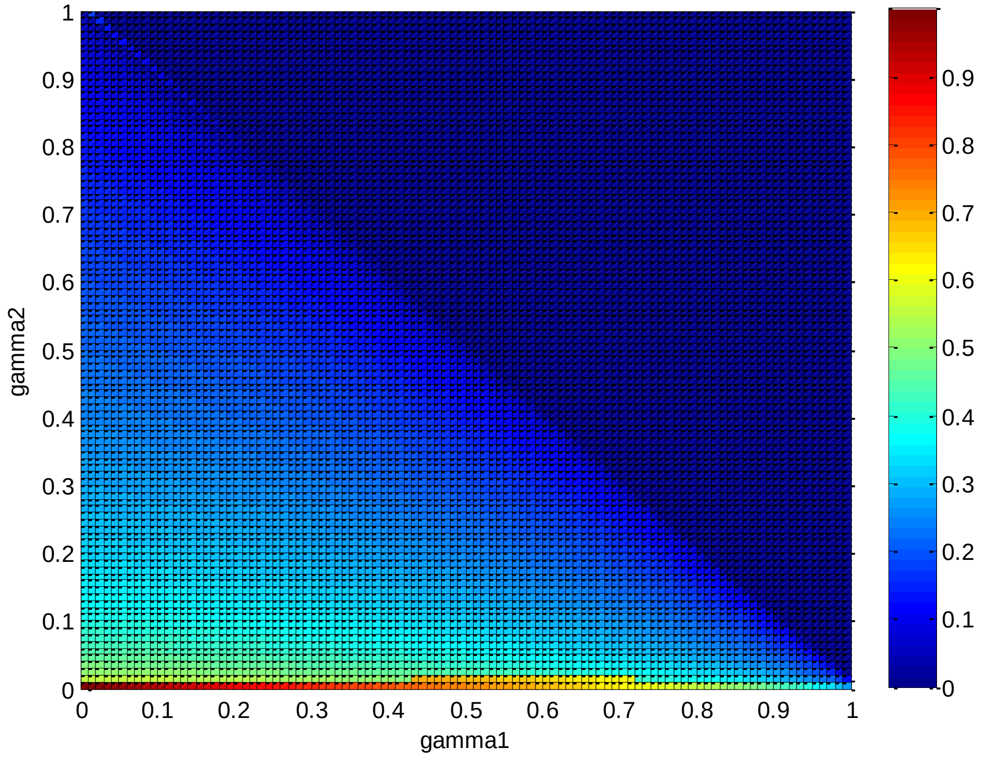


Figure 7c. Efficiency plot of loggene11 (projection to γ_1-γ_2 plane)

Although as mentioned earlier, equivalence between compound regression and constrained regression has not been theoretically proven, we can get some numerical results from this example. For example, if we set the constrained on γ_1 and γ_2 as follows: $e_1 \geq 0.6$ *and* $e_2 \geq 0.6$ we want to find the optimal γ_3 to maximize the e3.

It's easy to see that if we increase the proportion of γ_3, we can increase the efficiency of x3. We use Matlab code (see Appendix B.3) to calculate the corresponding e3. The output shows k=44, l=58 and the corresponding efficiency is 0.011.

We can try some other value. When we set criterion $e_1 \geq 0.3$ *and* $e_2 \geq 0.3$, we can get the optimal efficiency of x3 is 0.2207. When we set criterion $e_3 \geq 0.1$ *and* $e_2 \geq 0.3$, we can get the optimal efficiency of x1 is 0.9030.

4.4 Simulation Study

We have done analyses on real data in the previous section. Here we did some simulation study to test how well our model estimates the true slope.

2-dimension case:

We here conduct some simulation for test our model fitting. We create our data using the GLS model assumption.

Data generation

1) Data range, we define the range from 0 to 10.
2) ξ follows normal distribution with mean 0 and standard deviation 10;
3) Sample size, I choose 200 data points
4) True slope: $\eta=2\xi$.
5) Error term: error follows normal distribution with mean 0. Standard deviation of error in X (δ) is 0.1*range; standard deviation of error in Y (ε) is 0.1*range*abs(slope)

Model fitting

1) Compound regression:

We fit the data with our compound regression line and we have the following efficiency plot. (X-axis is γ, Y-axis is efficiency)

In this data set, we can see that range 0.15 to 0.25 is probably the best solution we need (in this situation, the efficiency of both X and Y is above 0.925).

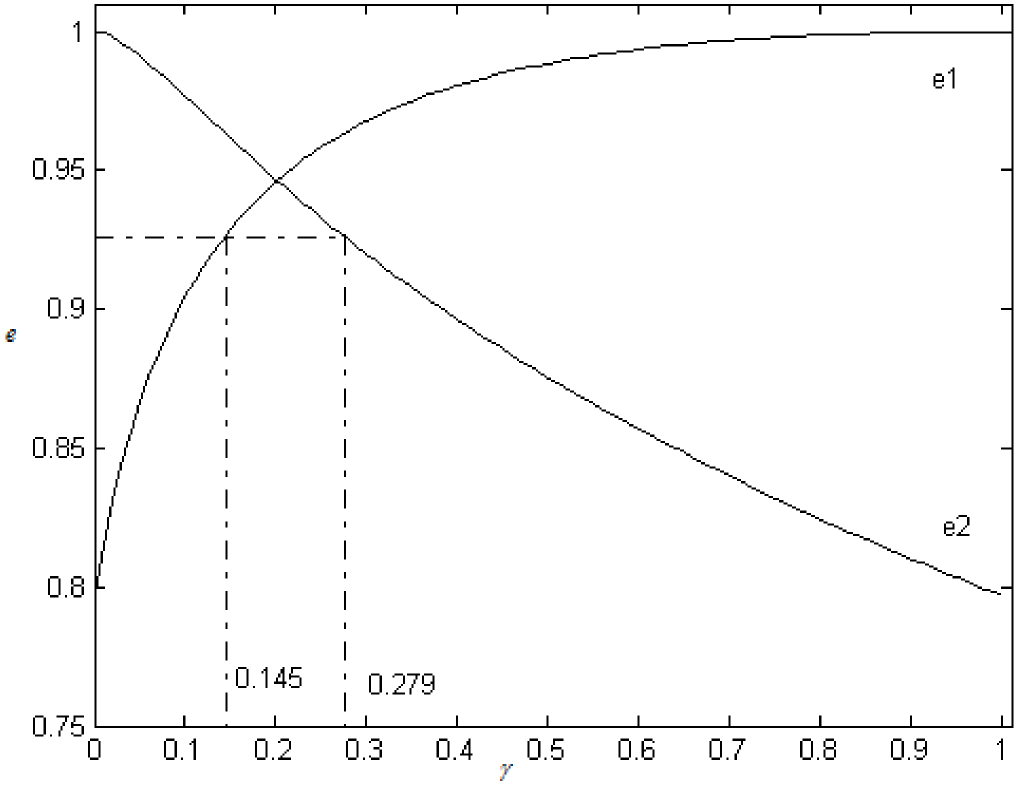


Figure 8a. Compound Regression Efficiency plot for simulation example1.

2) Constrained regression

From the plot, we can get the rough idea of obtaining the corresponding gamma to make sure the efficiency of regression for both variables would not be low. Now we run the constrained regression analyses to see in which range would both the efficiencies would be at least 0.925 and got the above results $\gamma \in [0.145, 0.279]$(in Figure 8a).

3) Generalized Least Square regression and other methods.

From the data generating step, we can see that GLS model assumptions hold here and the error variance ratio of Y and X is 4, therefore, the MLE of λ=4 should be suitable in this case and the corresponding $\beta = 1.9533$.

Compound and constrained regression give us the alternative selection method when we don't know the error ratio. In this example, γ corresponding to MLE of GLS equals 0.2155 which falls inside the interval we select. This means our model is close to the existing suitable model even when we have less information.

Re-sampling

Since we can't get theoretical inference of the slope estimate (confidence interval), we use re-sampling procedure to get the 95% confidence interval of slope estimate.

Procedure:

1) For each value of γ , randomly (with replacement) choose 200 points from the data set and calculate the corresponding slope estimate.
2) Do step 1 for 1000 times. Calculate the mean and standard deviation of β
3) Calculate the upper value as mean+1.96*std and lower value as mean-1.96*std.

4) Plot the confidence interval. X-axis is γ; Y-axis is the slope-estimate.

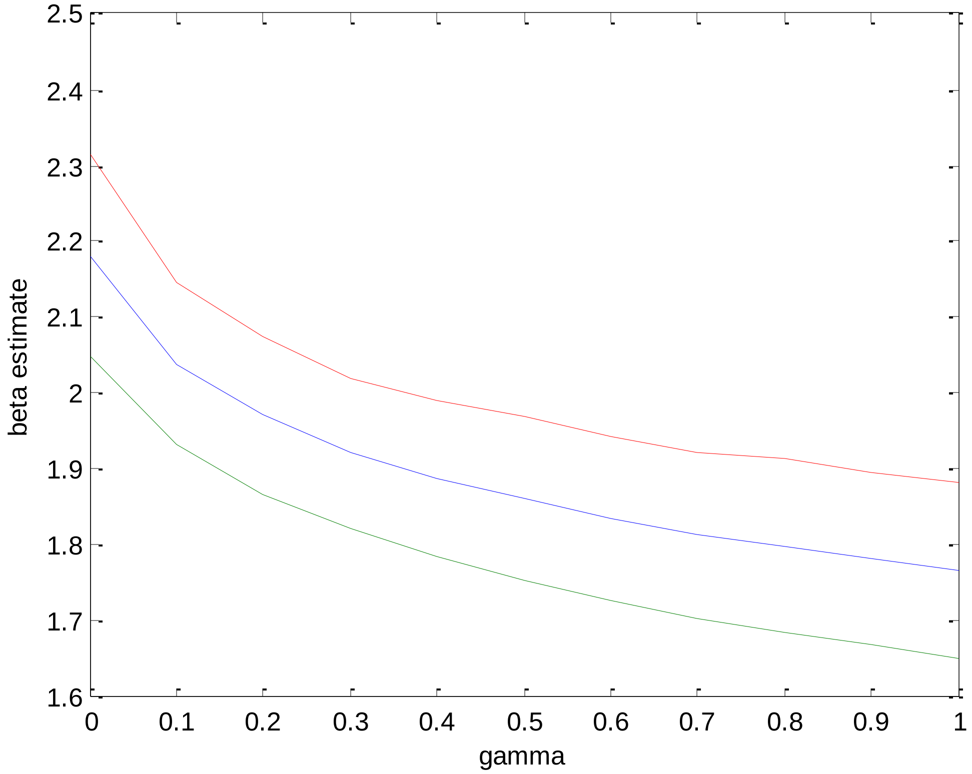


Figure 8b. Compound Regression confidence interval for simulation example1.

From the plot, we can see that the confidence interval of β when $\gamma \in [0.15, 0.28]$ covers the actual slope value of 2. Here, I also attached some slope estimate when $\gamma \in [0.15, 0.28]$.

$\gamma = 0.15,\ \ \beta = 2.0005 \qquad \gamma = 0.2,\ \ \beta = 1.9693$

$\gamma = 0.18,\ \ \beta = 1.9810 \qquad \gamma = 0.24,\ \ \beta = 1.9485$

Similarly, we can carry out experiments with different variance ratio, for example, let ratio equals 2, 1, 0.5 etc. Here we only list the results of these simulations.

1) Case: error variance ratio λ=2. In this case, we still use the model η=2ξ to generate data.

The efficiency plot is:

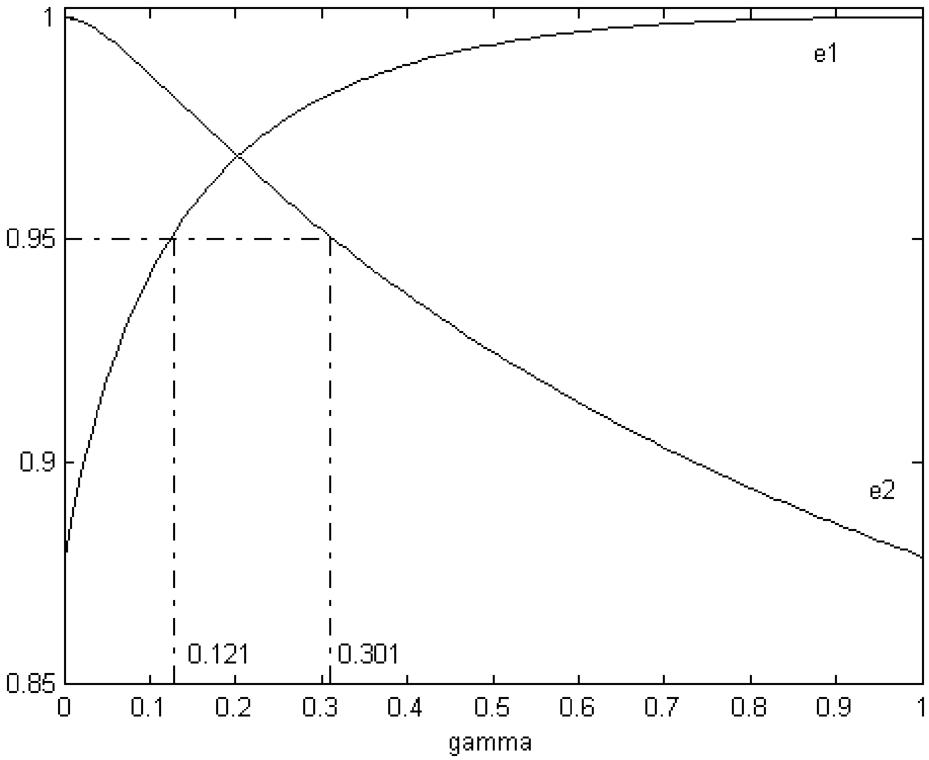


Figure 9a. Compound Regression Efficiency plot for simulation example2.

The MLE when $\lambda = 2$ is $\beta = 2.0286$. And from the efficiency plot above, we can see that if we want both efficiencies of regression to be no smaller than 0.95, we can limit the range of γ inside the interval of [0.121, 0.301]. This interval contains γ corresponding to the MLE (γ=0.1804). Here we also list the slope estimate corresponding to some γ value inside the interval.

$\gamma = 0.17,\ \beta = 2.0357$

$\gamma = 0.20,\ \beta = 2.0062$

$\gamma = 0.21,\ \beta = 1.9976$

A bootstrap re-sampling was carried out following the steps shown in simulation example1 and we got the following confidence interval plot.

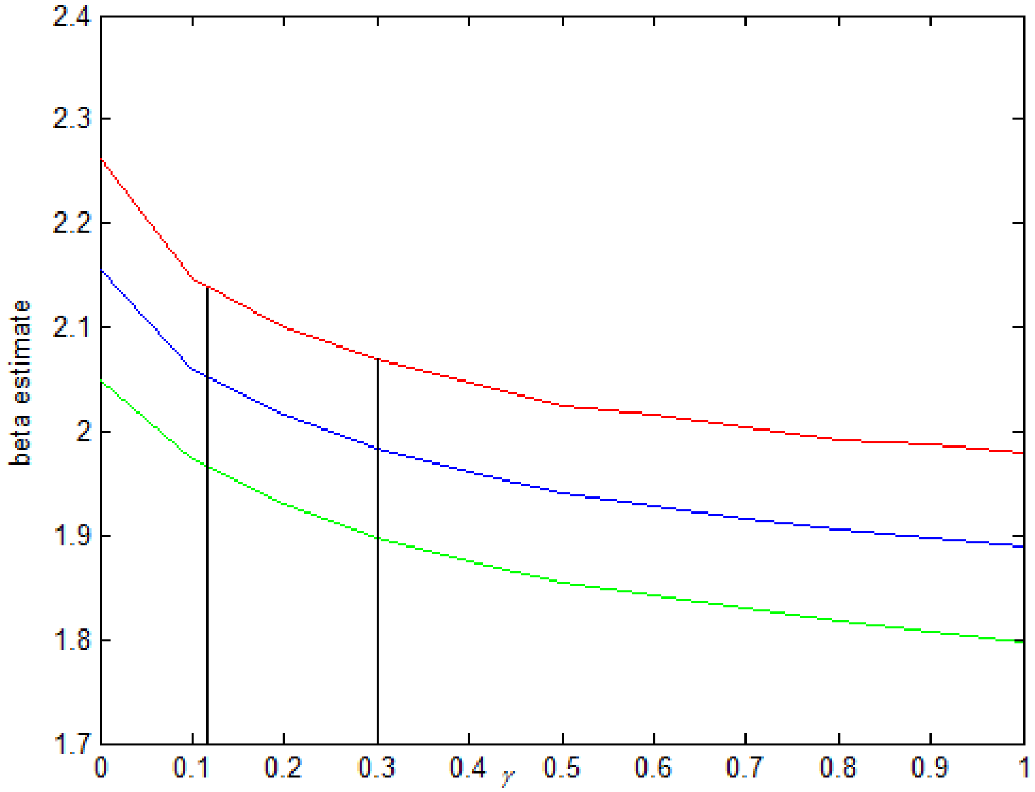


Figure 9b. Compound Regression confidence interval for simulation example2.

From the above plot, we can see that the true slope 2 is always inside the 95% confidence interval for the γ interval we select.

Case3: error variance ratio λ=1. This is the case when orthogonal regression is suitable.

The efficiency plot is:

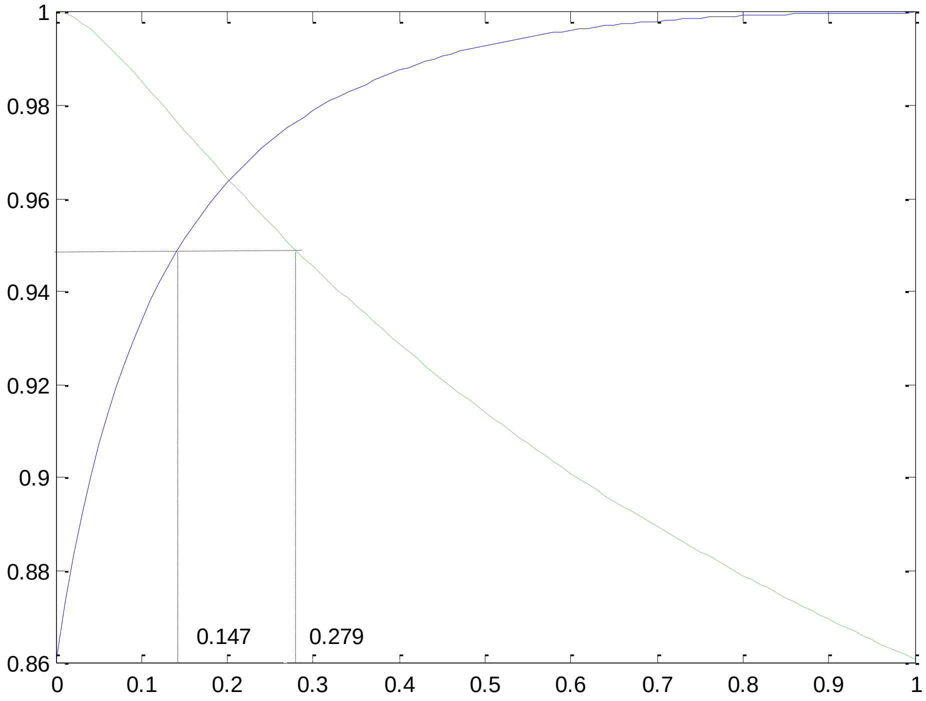


Figure 10. Compound Regression Efficiency plot for simulation example3.

From the efficiency plot, we can limit the range of γ from 0.147 to 0.279 then we can make both efficiencies larger than 0.95. The estimate from orthogonal regression is 2.0781 and the corresponding γ is 0.0509. Although this number is not inside the interval we select, we can see that the slope estimate of this interval varies from 2.0111 to 1.9577 which is closer to the exact slope parameter 2 we set in our simulation.

Unlike the GLS approach which requires the normality assumption, an advantage of our model is that it's non-parametric, that is, distribution-free. When we encounter data which does not follow normal distribution, GLS may not give us appropriate results. Here we give a simulation to illustrate this circumstance.

Data generation

2) Data range, we define the range from 0 to 10.

3) X follows *uniform* distribution with mean 0 and standard deviation 10;

4) Sample size, I choose 200 data points

5) True slope: η=2ξ

6) Error term: error follows *uniform* distribution with mean 0. Standard deviation of error in X (δ) is 0.2*range; standard deviation of error in Y(ε) is 0.2*range*sqrt(abs(slope))

And we have efficiency plot as follows:

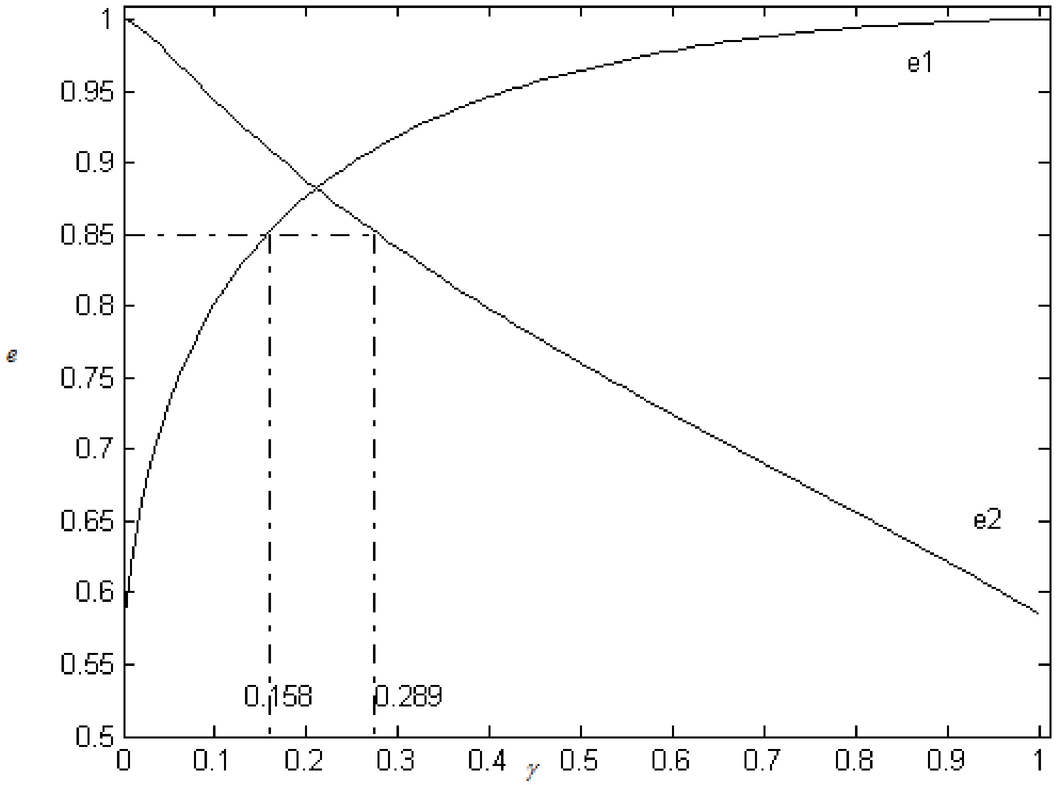


Figure 11a. Compound Regression Efficiency plot for simulation example4.

The interval we choose is [0.158, 0.289] in which efficiencies are no less than 0.85. The slope estimate we got in this interval varies from 1.861 to 2.0025 which are close to the exact value $\beta = 2$.

A bootstrap re-sampling was carried out and the 95% confidence interval plot is shown below:

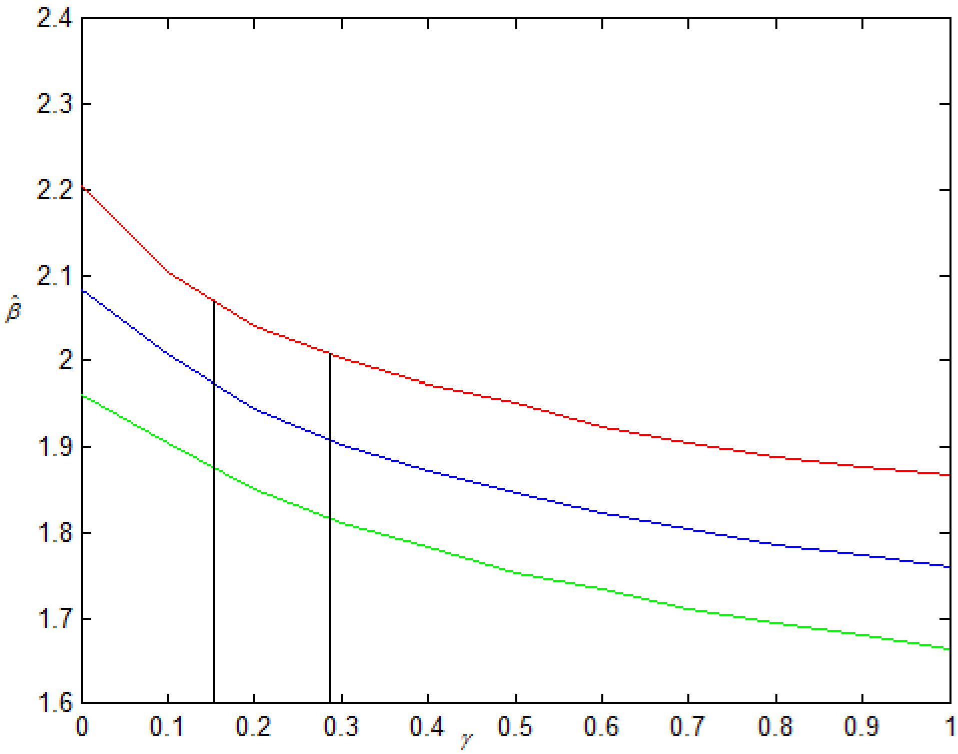


Figure11b. 95% confidence interval for slope estimate in simulation example 4

From the above plot, we can see that the true slope value 2 is always inside the 95% confidence interval for the β value in the interval we select.

We can compare this value with MLE. MLE of GLS is $\beta = 2.1242$ when in this case, λ=2. And orthogonal regression gives estimate of 2.2937. None of these estimates are as good as the compound and constrained regression estimate.

Here we give the scatter plot and regression lines.

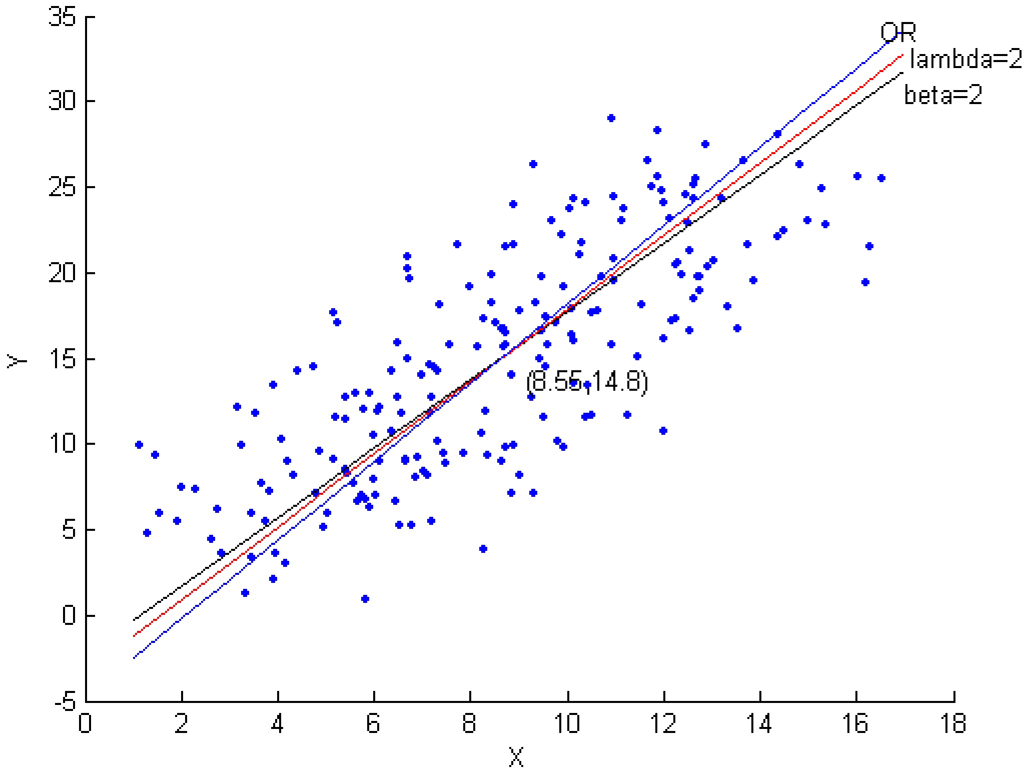


Figure 11c. Scatter plot for simulation example4.

We here also check our model when there are outliers in the dataset, now we still use the dataset above with several outlier points (2 points).

And we can see the scatter plot and efficiency plot below.

Table 4. Selected slope estimate Results for simulation Example4

Exact slope value	MLE(λ=2)	Orthogonal Regression	Compound Regression Interval	
2	2.2730	2.4998	2.0774	1.9024

From the above table, we can see that compound and constrained regression is the closest to the exact regression line.

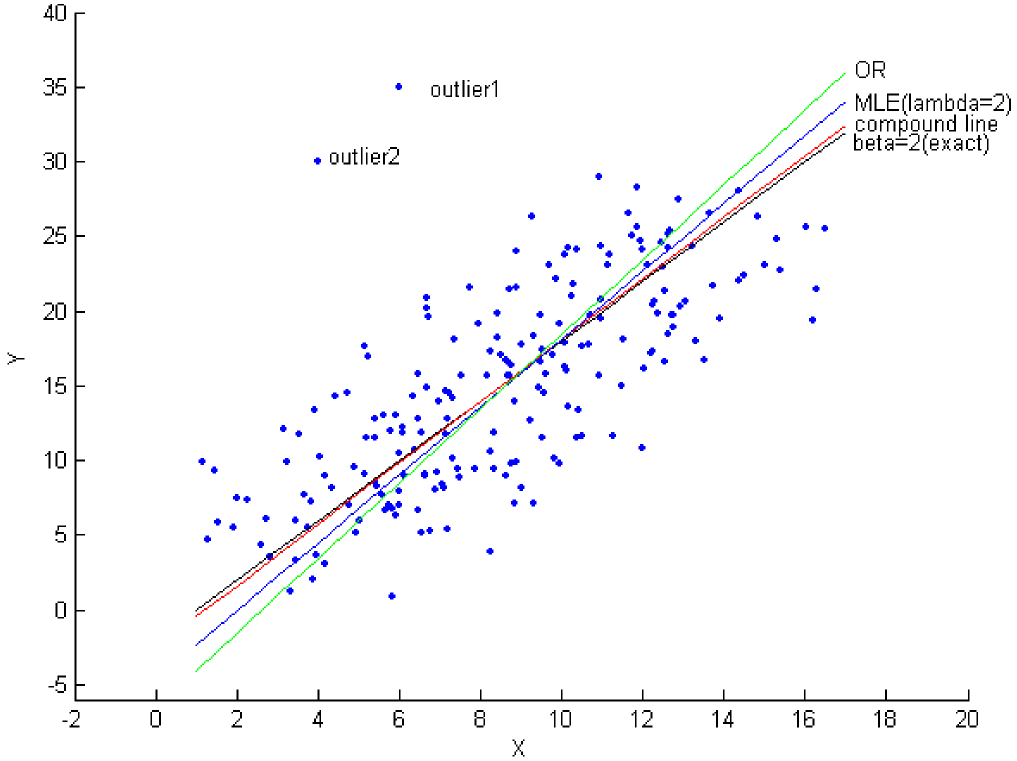


Figure 11d. Scatter plot for simulation example4 with outliers

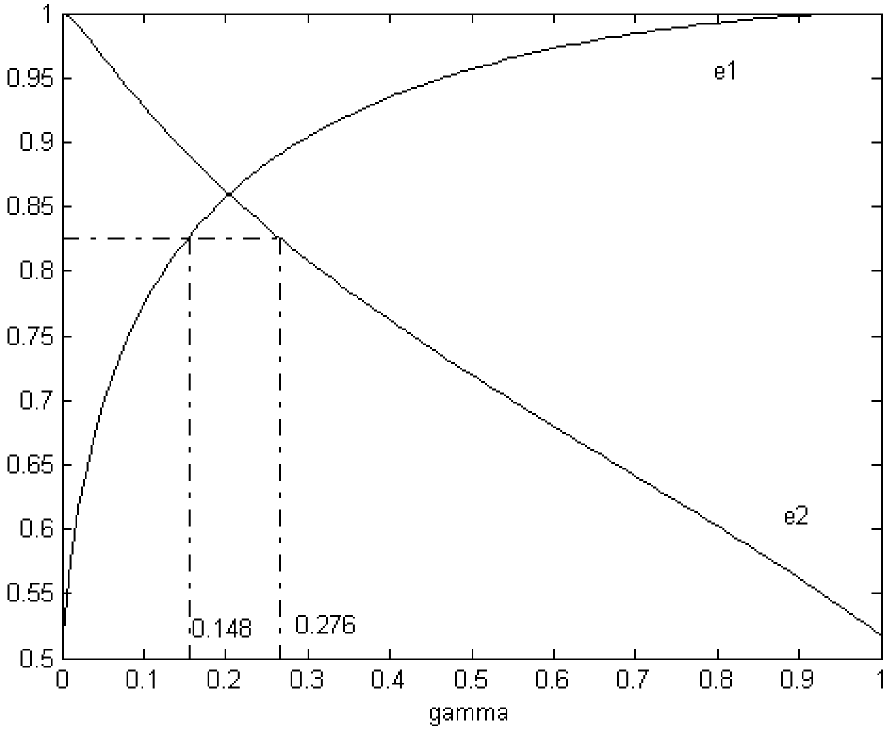


Figure 11e. Compound Regression Efficiency plot for simulation example4 with outliers

From this example, we can see that our model is non-parametric. Therefore, it has better performance when the normality assumption required by GLS can't be guaranteed.

Higher Dimension case

Now we check the performance of our model in higher dimension case.

Data generation

1) Data range, we define the range from 0 to 100. ξ_1 and ξ_2 follow normal distribution with mean 0, variance 100.

2) Sample size, I choose 200 data points.

3) True slopes follow the model: $\xi_3 = \xi_1 + \xi_2$.

4) Error term: variances of error in X (ε_1), Y (ε_2) and Z (ε_3) are 10.

Model fitting

1) Compound regression:

We fit the data with our compound regression model.

The efficiency plots of this data set are:

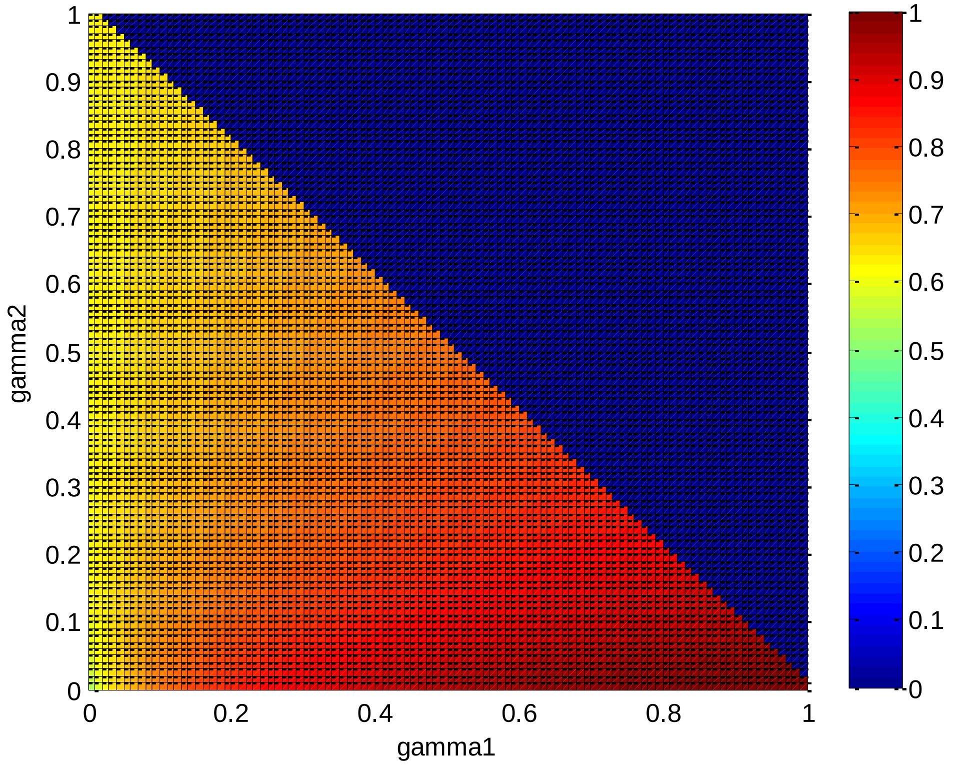


Figure 12a. Compound regression efficiency plot of Z (projection to γ_1-γ_2 plane) of simulation example 5

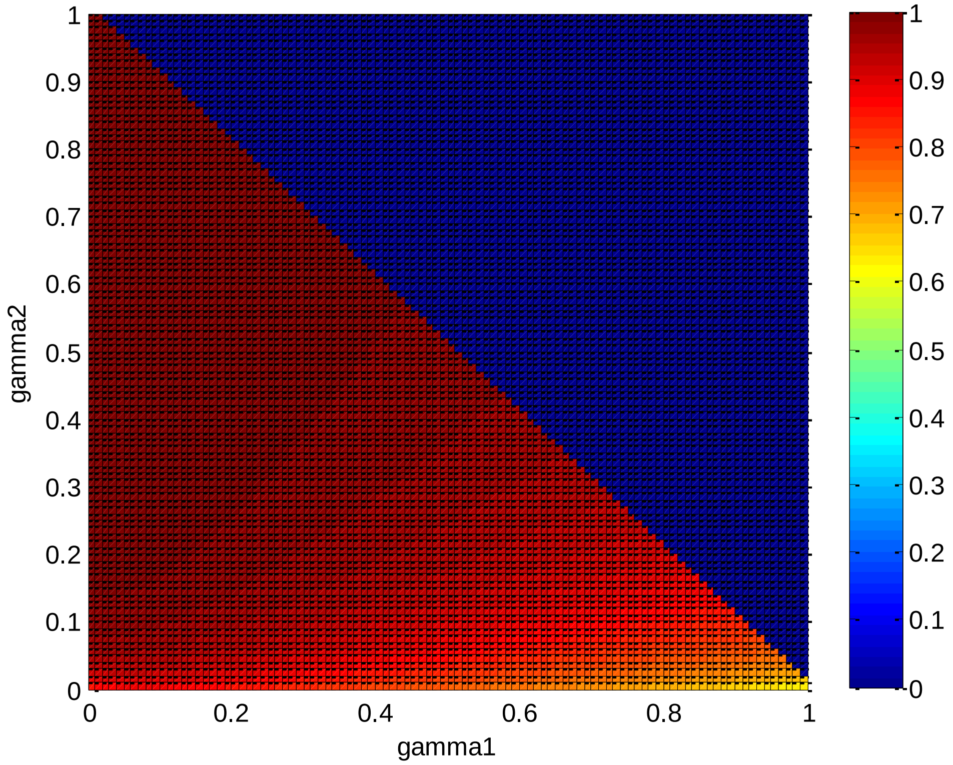


Figure 12b. Compound regression efficiency plot of X (projection to γ_1-γ_2 plane) of simulation example 5

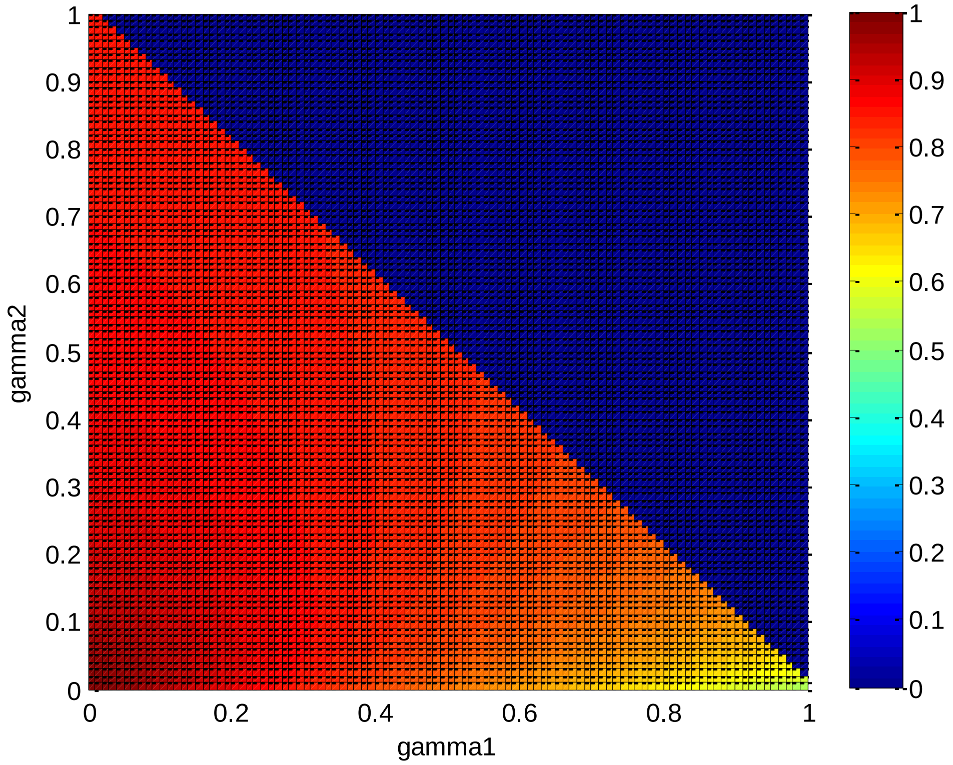


Figure 12c. Compound regression efficiency plot of Y (projection to γ_1-γ_2 plane) of simulation example 5

2) Constrained regression

Here we first use the criterion of efficiencies of regression of all the three variables be no less than 0.95. We use the following code:

```
c=0;

k=1

for i=1:101;

     for j=1:101;

          if e1(i,j)>=0.95 & e2(i,j)>=0.95 & e3(i,j)>=0.95;
```

```
                gamma1(k)=i;
                gamma2(k)=j;
                k=k+1;
            else;
            end;
        end;
end;
```

And we have γ_1 and γ_2 contain the combination of γ satisfying our criterion. There are 687 points satisfy this criterion, that is, about7%. And the slope estimate varies from 0.9793 to 1.0113 for β_1 and from 1.0052 to 1.0385 for β_2. These values are in a small range around the exact value of (1, 1).

3) Generalized Least Square Regression

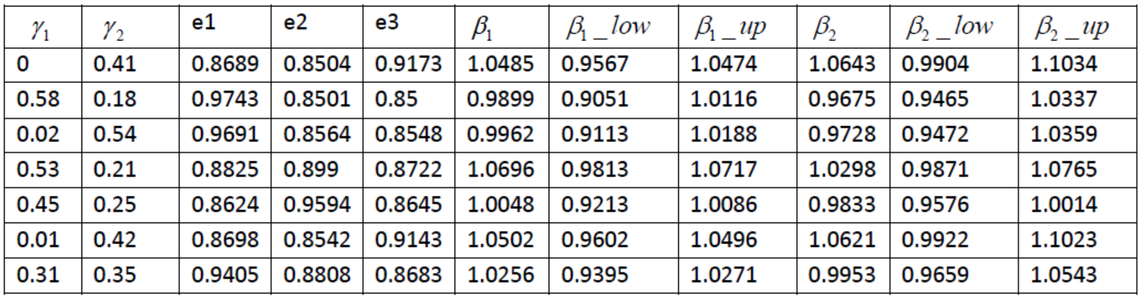
The MLE of GLS here is β_1=0.9878, β_2=1.0161 when Λ=I_3. And the corresponding γs are γ_1=0.3313, γ_2=0.3155 and efficiency of regressions are e_1=0.9786, e_2=0.9565, e_3=0.9615. This shows that MLE falls in the interval we select.

4) Re-sampling

Here we did a bootstrap re-sampling to obtain the confidence interval for the slope estimate, and we found that the true slope 1 is always inside the 95% confidence intervals.

We have the following table containing selected value.

Table 5. selected output of 95% confidence interval from re-sampling



γ_1	γ_2	e1	e2	e3	β_1	β_1_low	β_1_up	β_2	β_2_low	β_2_up
0	0.41	0.8689	0.8504	0.9173	1.0485	0.9567	1.0474	1.0643	0.9904	1.1034
0.58	0.18	0.9743	0.8501	0.85	0.9899	0.9051	1.0116	0.9675	0.9465	1.0337
0.02	0.54	0.9691	0.8564	0.8548	0.9962	0.9113	1.0188	0.9728	0.9472	1.0359
0.53	0.21	0.8825	0.899	0.8722	1.0696	0.9813	1.0717	1.0298	0.9871	1.0765
0.45	0.25	0.8624	0.9594	0.8645	1.0048	0.9213	1.0086	0.9833	0.9576	1.0014
0.01	0.42	0.8698	0.8542	0.9143	1.0502	0.9602	1.0496	1.0621	0.9922	1.1023
0.31	0.35	0.9405	0.8808	0.8683	1.0256	0.9395	1.0271	0.9953	0.9659	1.0543

We can also get the simulation on different λs. Here we list some results:

Case2: We are using error variance of X 20, error variance of Y 40, error variance of Z 10.

That is,

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Then we have the following results.

Efficiency plots

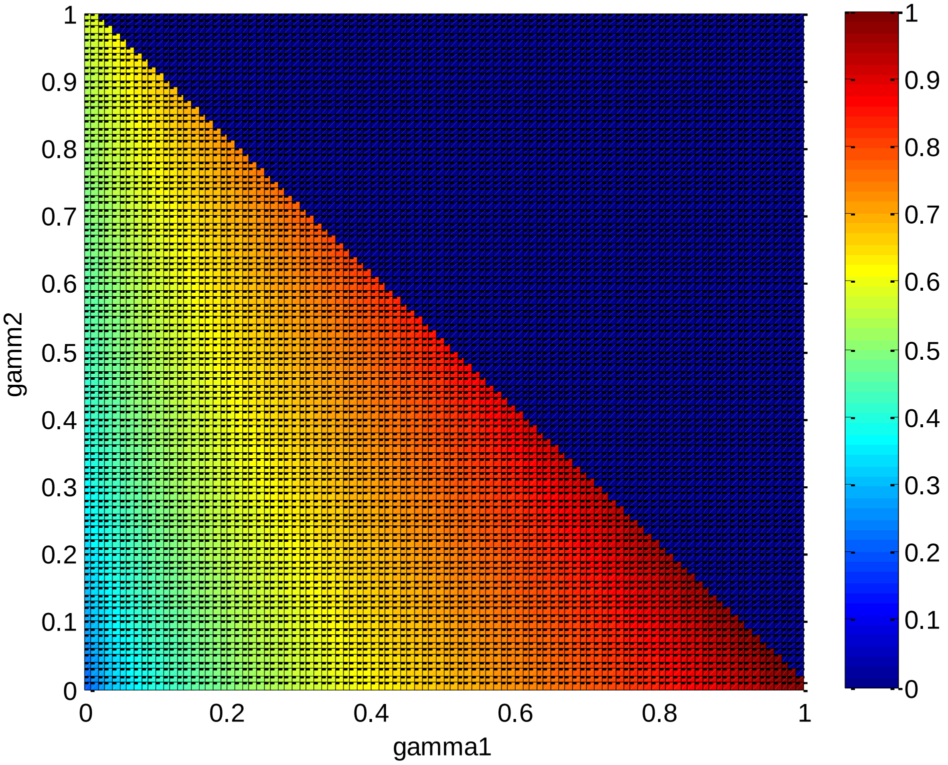


Figure 13a. Compound regression efficiency plot of Z (projection to γ_1-γ_2 plane) of simulation example 6

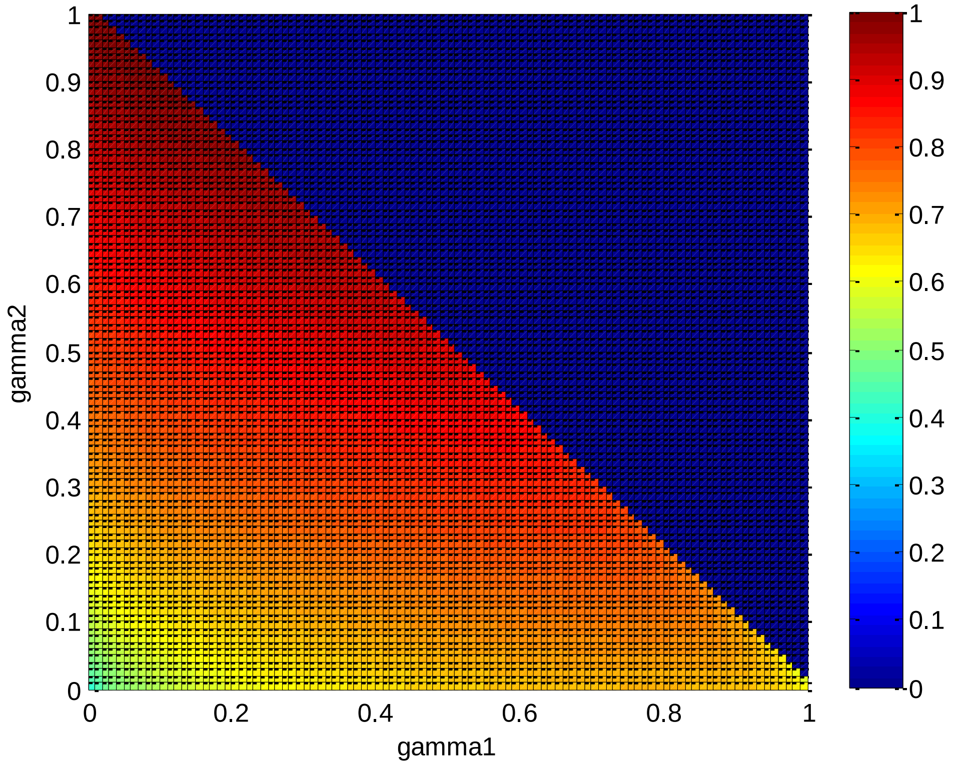


Figure 13b. Compound regression efficiency plot of X (projection to γ_1-γ_2 plane) of simulation example 6

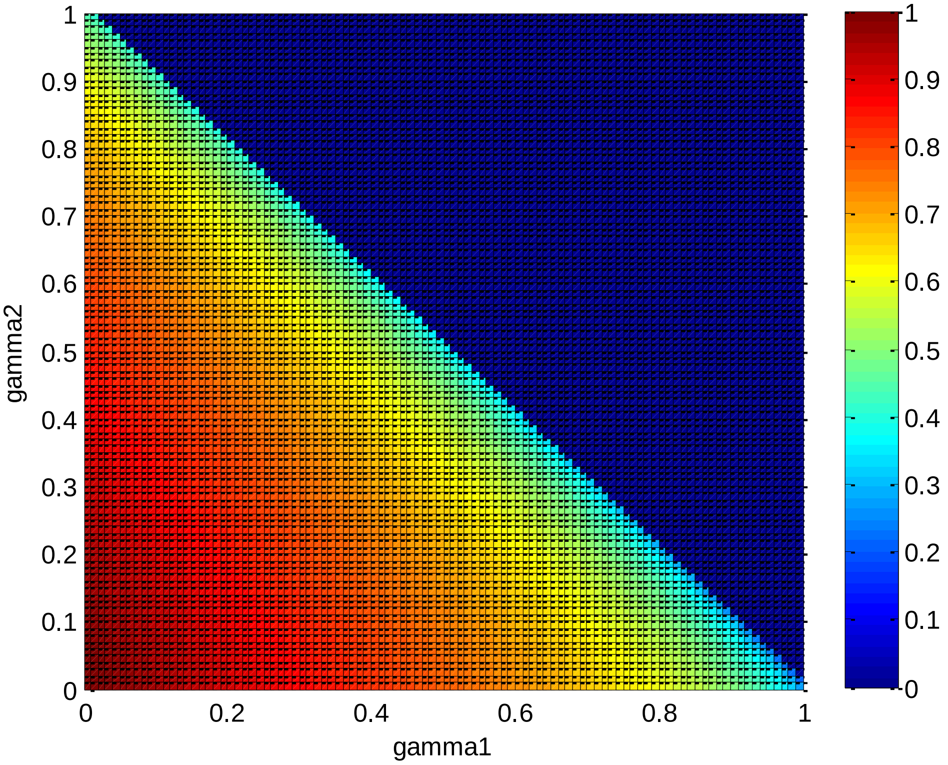


Figure 13c. Compound regression efficiency plot of Y (projection to γ_1-γ_2 plane) of simulation example 6

In order to select the interval, we set the criterion to efficiencies no less than 0.75. Then we can see that the estimate of β_1 varies from 0.9645 to 1.0407 and the estimate of β_2 varies from 0.9345 to 1.0121. The MLE of GLS is β_1=1.0158, β_2=0.9449 with corresponding γ_1=0.1562, γ_2=0.3459 and efficiencies are e_1=0.7734, e_2=0.7434, e_3=0.7927. We can see that this clearly fall into the interval we are selecting.

Case 3, we simulate different model. In this case, we are using model: $\xi_3=2\xi_1+3\xi_2$.

Data generation

1) Data range, we define the range from 0 to 10. ξ_1 and ξ_2 follows normal distribution with mean 0 and standard derivation of 10;

2) Sample size, I choose 200 data points

3) True slopes follow the model: $\xi_3=2\xi_1+ 3\xi_{2.}$

4) Error term: variance of error in X (ε_1) is 20; variance of error in Y (ε_2) is 40 and variance of error in Z (ε_3) is 10.

Model fitting

1) Compound regression

We have the following efficiency plots:

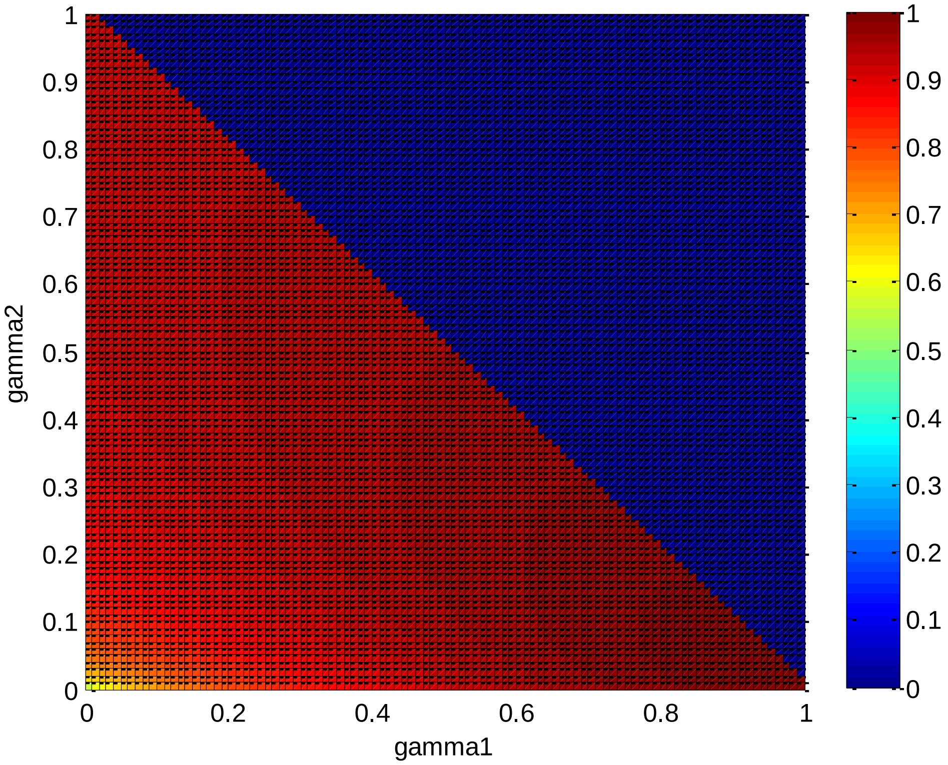


Figure 14a. Compound regression efficiency plot of Z (projection to γ_1-γ_2 plane) of simulation example 7

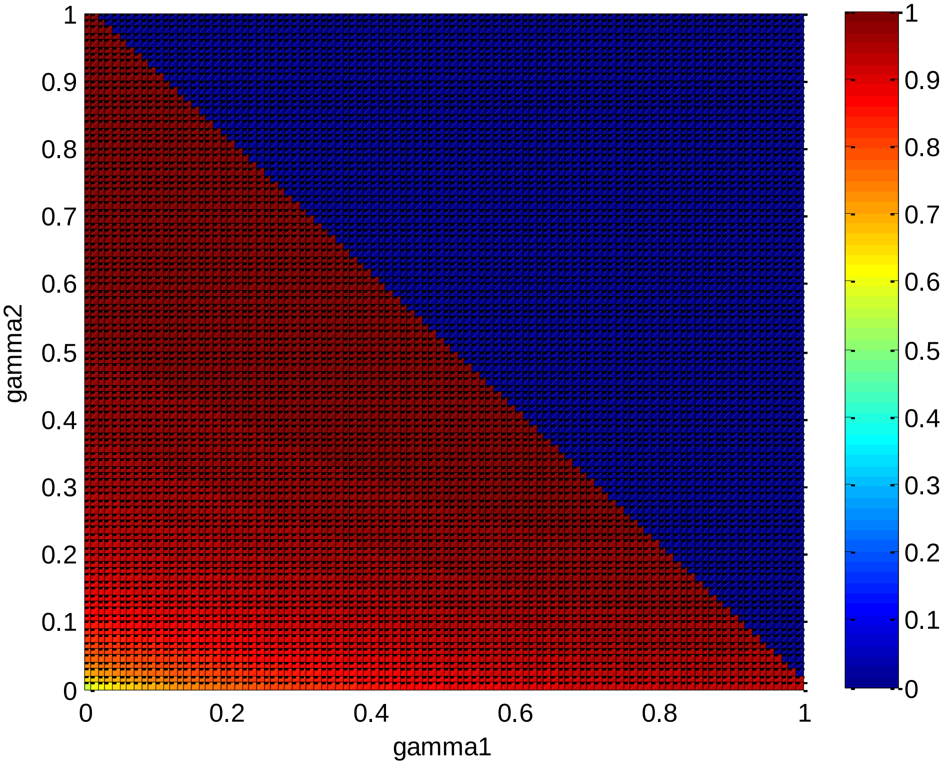


Figure 14b. Compound regression efficiency plot of X (projection to γ_1-γ_2 plane) of simulation example 7

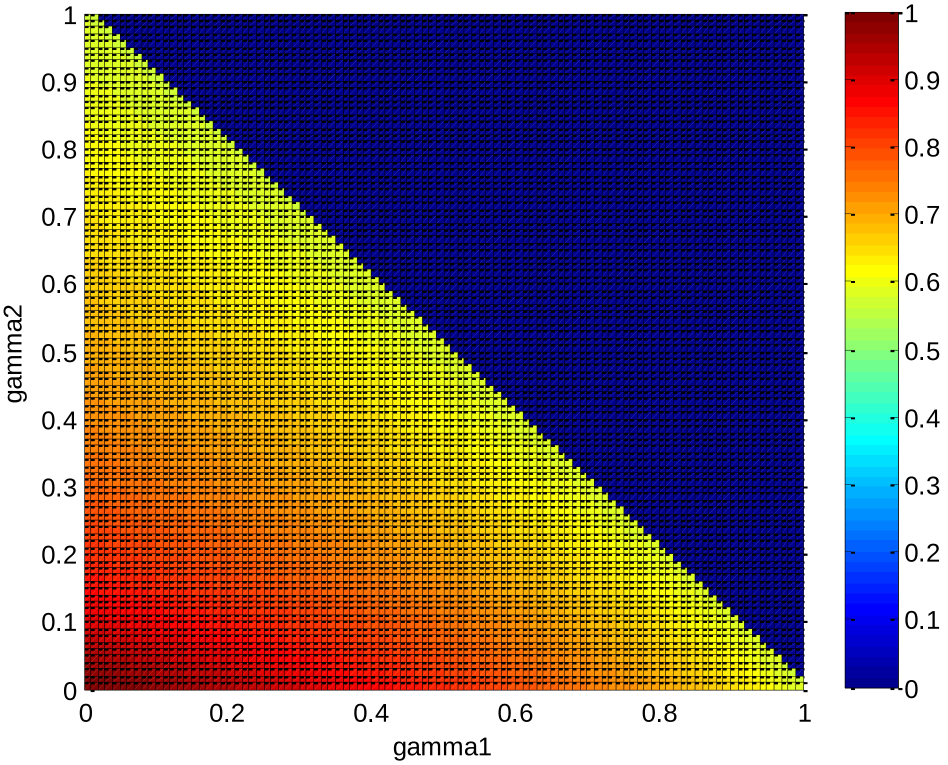


Figure 14c. Compound regression efficiency plot of Y (projection to γ_1-γ_2 plane) of simulation example 7

And we set the efficiency criterion to be no less than 0.8. The estimate of β_1 varies from 2.0129 to 2.1226 and the estimate of β_2 varies from 2.8932 to 3.1106.

MLE of GLS gives us the slope estimates are $\hat{\beta}_1$=2.0825, $\hat{\beta}_2$=3.1849 and the corresponding γs are γ_1=0.0279, γ_2=0.4452. And the efficiencies of regression are: e_1=0.8119, e_2=0.8037, e_3=0.8351.

From the efficiency value, we can see that MLE falls inside the interval we select.

In summary, simulation shows that our compound and constrained analyses, along with the definition of efficiency, give us an alternative way to select the more suitable regression line. In circumstances that the ratio of variance is not available, we can still get very close fit to the real slope and the MLE of GLS.

5. DISCUSSION

In this work, we presented two novel approaches for deriving the best regression line for regression with errors in variables (EIV), also known as the measurement error model. We derived the equivalence between the compound regression approach and the traditional maximum likelihood estimation (MLE) method for generalized linear model. Furthermore, we proved the equivalence between the compound and the constrained regression approaches analytically for the simple linear regression model while demonstrated their equivalency in higher dimensions numerically.

Our approaches are distribution free while the traditional Frequentist MLE method relies on the multivariate normality assumption (Kerridge 1967). Statistical inference based on the compound regression analysis or equivalently, the constrained regression analysis approach can be carried out using the nonparametric bootstrap resampling method (Efron 1979; Efron and Tibshirani 1993).

In this thesis, we focused on the Frequentist approach due to the lack of prior knowledge/information in most situations. We will examine potential extension of our approaches in a Bayesian context for the future and compare them to existing Bayesian methods (Zellner 1971; Bretthort, 1988). In addition, we will also compare the performance of our methods to existing robust regression methods such as the least median-of-squares method (Rousseeuw, 1984) which represent an alternative direction of development for EIV models.

REFERENCES

Barker, F., Soh,Y. C., and Evans, R. J. (1988), “Properties of the Geometric Mean Functional Relationship,” *Biometrics*, 44, 279-281.

Biedermann, S., Dette, H., and Zhu, W. (2006),“Optimal Designs for Dose-Response Models with Restricted Design Spaces,” *Journal of the American Statistical Association*, 101, 747 -759.

Bretthorst, G. L. (1988), “Bayesian Spectrum Analysis and Parameter Estimation.” Lecture notes in statistics, Vol. 48. Springer-Verlag, Berlin.

Casella, G., and Berger, R. L. (2001), “Statistical Inference (Second Edition)”. Duxbury.

Cook, R. D., and Wong, W. K. (1994), “On the Equivalence of Constrained and Compound Optimal Designs,” *Journal of the American Statistical Association,* 89, 687-692

Clyde, M., and Chaloner, K. (1996), “The Equivalence of Constrained and Weighted Designs in Multiple Objective Design Problems,” *Journal of the American Statistical Association*, 91, 1236 -1244.

Dette, H., Wong, W.K., and Zhu, W. (2005), “On the Equivalence of Optimality Design Criteria for the Placebo-Treatment Problem,” *Statistics and Probability Letters*, 74, 337-346.

Efron, B. (1979), “Bootstrap Methods: Another Look at the Jackknife,” *Annals of Statistics*, 7, 1-26.

Efron, B., and Tibshirani, R. (1993), “An Introduction To the Bootstrap,” *New York: Chapman & Hall/CRC.*

J. D. Jackson and J. A. Dunlevy.(1988), “Orthogonal Least Squares and the Interchangeability of Alternative Proxy Variables in the Social Sciences,” *The Statistician*, 37(1): 7-14, 1988.

Jaynes, E. T. (2004), “Probability Theory – The logic of science”. Cambridge University Press, Cambridge, UK, 2004.

Jobson, B. T., Mckeen, S. A, Parrish, D.D., Fehsenfeld, F. C., Blake, D. R., Goldstein, A. H., Schauffler, S. M., and Elkins, J. W. (1999), “Trace Gas Mixing Ratio Variablility Versus Lifetime in the Troposphere and Stratosphere: Observations,” *Journal of Geophysical Research*, 104, 16091-16113

Kerridge, D. (1967), “Errors of Prediction in Multiple Regression with Stochastic Regressior Variables,” *Technometrics*, 9, 309-311.

Kleinman, L. I., Springston, S. R., Daum, P. H., Lee, Y.-N., Nunnermacker, L. J., Senum, G.I., Wang, J., Weinstein-Lloyd, J., Alexander, M. L., Hubbe, J., Ortega, J., Canagaratna, M. R., and Jayne, J. (2007), “The Time Evolution of Aerosol Composition over the Mexico City Plateau,” *Atmospheric Chemistry and Physics*, 7, 1-49.

Lindley, D. V. (1947) *Supplement to the Journal of the Royal Statistical Society*, 9, 218-244.

Läuter, E. (1974), “Experimental Planning in A Class of Models,” *Mathematische*

Operationsforschung Und Statistik, 5, 673-708.

Läuter, E. (1976), "Optimal Multipurpose Designs for Regression Models," M*athematische Operationsforschung Und Statistik*, 7, 1-68.

Lee, C. M. S. (1987), "Constrained Optimal Designs for Regression Models," *Communications in Statistics*, Part A-Theory and Methods, 16, 765-783.

Lee, C. M. S. (1988), "Constrained Optimal Design," *Journal of Statistical Planning and Inference*, 18, 377-389.

Mikulecka, J., Jackson, J. D., and Dunlevy, J. A. (1983), "Orthogonal Least Squares and the Interchangeability of Alternative Proxy Variables in the Social Science," *The Statistician*, 37, 7-14, 1988.

Patefield, W. M. (1981), "Multivariate Linear Relationships: Maximum Likelihood Estimation and Regression Bounds," Journal of Royal Statistics Society B, 43, 342-352.

Rousseeuw P. J. (1984), "Least Median of Squares Regression." *Journal of the American Statistical Association*, 79, 871–880.

Sprent, P., and Dolby, G. R. (1980), "Query: the Geometric Mean Functional Relationship," *Biometrics*, 36, 547-550.

Sprent, P. (1969), "Models in Regression and Related Topics," *Methuen's Statistical Monographs.*

Wong, M.Y. (1989), "Likelihood Estimation of A Simple Linear Regression Model When Both Variables Have Error," *Biometrica*, 76, 141-148

Zellner, A. (1987), “An Introduction to Bayesian Inference in Econometrics (2nd edition).” R.E. Krieger Pub. Co., Malabar, Florida.

Zhu, W. (1996), “On the Optimal Designs of Multiple-Objective Clinical Trials and Quantal Dose Response Experiments,” UCLA School of Public Health, PhD Thesis.

Zhu, W., and Wong, W. K. (1998), “Multiple-Objective Designs in Dose-Response Experiments.” Institute of Mathematical Statistics Lecture Notes --Monograph Series: *New Developments and Applications in Experimental Designs*, 73-82.

APPENDIX A: PROOFS OF RESULTS

Theorem 1: the orthogonal regression is the same as the MLE when all the errors are equal ($\lambda = \mathbf{1}$).

Proof:

The only thing we care is the slope estimates (not the intercept term), so it would be OK if we just consider the case of centered data.

For the MLE approach, the slope estimate is the eigenvector corresponding to the smallest eigenvalue of matrix S (the covariance matrix). This is actually very reasonable. Since the covariance matrix is nonsingular, and the PCs can expand the same dimension (n) of space of data and the regression plane is (n-1) dimension. Hence, it is reasonable to use only the PCs that can explain most variance to expand the regression plane. we can define these eigenvectors as $\gamma_1, \gamma_2, ..., \gamma_k$ in the order of descending eigenvalues ($c_1, c_2, ..., c_k$).

In orthogonal regression under the above assumption, we can see that the formula has minimizing $\sum_{i=1}^{n} \frac{(\beta_1 X_{1i} + \beta_2 X_{2i} + ...\beta_k X_{ki})^2}{\beta_1^2 + \beta_2^2 + ... + \beta_k^2}$. We can put the above formula in matrix form which is

$$\sum_{i=1}^{n} (\frac{\beta_1}{\sum \beta_j^2}, \frac{\beta_2}{\sum \beta_j^2}, ..., \frac{\beta_k}{\sum \beta_j^2})(x_{i1}, x_{i2}, ..., x_{ik})^T (x_{i1}, x_{i2}, ..., x_{ik})(\frac{\beta_1}{\sum \beta_j^2}, \frac{\beta_2}{\sum \beta_j^2}, ..., \frac{\beta_k}{\sum \beta_j^2})^T$$

$$=\sum_{i=1}^{n}(\frac{\beta_1}{\sum\beta_j^2},\frac{\beta_2}{\sum\beta_j^2},...,\frac{\beta_k}{\sum\beta_j^2})\begin{bmatrix} x_{i1}^2 & ... & x_{i1}x_{ik} \\ ... & ... & ... \\ x_{i1}x_{ik} & ... & x_{ik}^2 \end{bmatrix}(\frac{\beta_1}{\sum\beta_j^2},\frac{\beta_2}{\sum\beta_j^2},...,\frac{\beta_k}{\sum\beta_j^2})^T$$

$$=(\frac{\beta_1}{\sum\beta_j^2},\frac{\beta_2}{\sum\beta_j^2},...,\frac{\beta_k}{\sum\beta_j^2})\sum_{i=1}^{n}\begin{bmatrix} x_{i1}^2 & ... & x_{i1}x_{ik} \\ ... & ... & ... \\ x_{i1}x_{ik} & ... & x_{ik}^2 \end{bmatrix}(\frac{\beta_1}{\sum\beta_j^2},\frac{\beta_2}{\sum\beta_j^2},...,\frac{\beta_k}{\sum\beta_j^2})^T$$

$$=(\frac{\beta_1}{\sum\beta_j^2},\frac{\beta_2}{\sum\beta_j^2},...,\frac{\beta_k}{\sum\beta_j^2})\begin{bmatrix} \sum_{i=1}^{n}x_{i1}^2 & ... & \sum_{i=1}^{n}x_{i1}x_{ik} \\ ... & ... & ... \\ \sum_{i=1}^{n}x_{i1}x_{ik} & ... & \sum_{i=1}^{n}x_{ik}^2 \end{bmatrix}(\frac{\beta_1}{\sum\beta_j^2},\frac{\beta_2}{\sum\beta_j^2},...,\frac{\beta_k}{\sum\beta_j^2})^T$$

$$=(\frac{\beta_1}{\sum\beta_j^2},\frac{\beta_2}{\sum\beta_j^2},...,\frac{\beta_k}{\sum\beta_j^2})S(\frac{\beta_1}{\sum\beta_j^2},\frac{\beta_2}{\sum\beta_j^2},...,\frac{\beta_k}{\sum\beta_j^2})^T$$

where S is the covariance matrix.

We can see that these β s are normalized vectors.

Since the eigenvectors can expand the space. We can write the above vector in $l_1\gamma_1+l_2\gamma_2+..+l_k\gamma_k$ $(l_1+l_2+...+l_k=1)$ therefore, the above formula will become:

$$(l_1\gamma_1+l_2\gamma_2+..+l_k\gamma_k)S(l_1\gamma_1+l_2\gamma_2+..+l_k\gamma_k)^T$$

And we know the eigenvectors are orthogonal and $\gamma_i S\gamma_i^T=c_i$, therefore, the above formula becomes $c_1l_1+c_2l_2+...+c_kl_k$. Under the constraint $l_1+l_2+...+l_k=1$ and $c_1\ge c_2\ge...\ge c_k$, we can obtain the minimum when $l_k=1$ which is the eigenvector corresponding to the smallest eigenvalue. That is, the MLE. □

Theorem 3 Equivalence of the constrained regression and the compound regression in 2-dimensional case.

Proof. The proof is divided into three parts. The first two parts are

1) $\forall\gamma$, suppose the estimator minimizing $(1-\gamma)\sum_{i=1}^{n}(X_i-\tilde{X}_i)^2+\gamma\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2$ is $\hat{\beta}_\gamma$, then there exists a value c such that $\hat{\beta}_\gamma$ will also minimize $\sum_{i=1}^{n}(X_i-\tilde{X}_i)^2$ under the constraint $\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2\le c$.

Proof. $\forall\gamma$, suppose $\hat{\beta}_\gamma$ will minimize $(1-\gamma)\sum_{i=1}^{n}(X_i-\tilde{X}_i)^2+\gamma\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2$,

Let $c=S_{YY}(\hat{\beta}_\gamma)\square\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2|_{\beta=\hat{\beta}_\gamma}$, $\hat{\beta}_\gamma$ will minimize $\sum_{i=1}^{n}(X_i-\tilde{X}_i)^2$ under the constraint that $\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2\le c$; otherwise, we will have $\hat{\beta}'$ satisfy: (1) $S_{YY}(\hat{\beta}')\le c$; (2) $S_{XX}(\hat{\beta}')\le S_{XX}(\hat{\beta}_\gamma)$ which yields

$$\gamma S_{YY}(\hat{\beta}')+(1-\gamma)S_{XX}(\hat{\beta}')\le\gamma c+(1-\gamma)S_{XX}(\hat{\beta}_\gamma)\le\gamma S_{YY}(\hat{\beta}_\gamma)+(1-\gamma)S_{XX}(\hat{\beta}_\gamma)$$

This contradicts the fact that "$\hat{\beta}_\gamma$ would minimize $(1-\gamma)\sum_{i=1}^{n}(X_i-\tilde{X}_i)^2+\gamma\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2$"

Therefore, there exists a value c for which $\hat{\beta}_\gamma$ would minimize the constrained regression model.

2) $\forall c$, under the constraint of $\sum_{i=1}^{n}(Y_i-\tilde{Y}_i)^2\le c$, suppose the estimator minimizing

$\sum_{i=1}^{n}(X_i - \tilde{X}_i)^2$ is $\hat{\beta}_c$, then there exists a γ, such that $\hat{\beta}_c$ is also the estimator to minimize $(1-\gamma)\sum_{i=1}^{n}(X_i - \tilde{X}_i)^2 + \gamma\sum_{i=1}^{n}(Y_i - \tilde{Y}_i)^2$.

Proof. First, we prove that $S_{YY}(\beta_\gamma)$ is a continuous and monotonic function of γ . Here we use the inverse function of Equation (1):

$$\frac{\gamma}{1-\gamma}\beta_1^4 S_{XX} - \frac{\gamma}{1-\gamma}\beta_1^3 S_{XY} + \beta_1 S_{XY} - S_{YY} = 0$$

Since β is monotone in γ when γ varies from 0 to 1 (Theorem 1b), the inverse function should exist and γ can be solved from

$$k = \frac{\gamma}{1-\gamma} = \frac{1}{\beta_1^3}\frac{S_{YY} - S_{XY}\beta_1}{S_{XX}\beta_1 - S_{XX}}$$

The right part is a continuous function and thus the inverse function when γ varies from 0 to 1 is also continuous. Therefore β is continuous of γ , and we know that $S_{YY}(\beta)$ is a continuous function of β which obtain that $S_{YY}(\beta_\gamma)$ is a continuous of γ .

Let $\gamma_1 > \gamma_2$ since β_{γ_1} will minimize the compound function for $\gamma = \gamma_1$ we have:

$$\begin{aligned}&(1-\gamma_1)S_{XX}(\beta_{\gamma_1}) + \gamma_1 S_{YY}(\beta_{\gamma_1}) \le (1-\gamma_1)S_{XX}(\beta_{\gamma_2}) + \gamma_1 S_{YY}(\beta_{\gamma_2})\\ &\Rightarrow (1-\gamma_1)(S_{XX}(\beta_{\gamma_1}) - S_{XX}(\beta_{\gamma_2})) + \gamma_1(S_{YY}(\beta_{\gamma_1}) - S_{YY}(\beta_{\gamma_2})) \le 0\end{aligned} \quad \text{(A1)}$$

Similarly, since β_{γ_2} will minimize the compound function when $\gamma = \gamma_2$, we have:

$$(1-\gamma_2)(S_{XX}(\beta_{\gamma_2}) - S_{XX}(\beta_{\gamma_1})) + \gamma_2(S_{YY}(\beta_{\gamma_2}) - S_{YY}(\beta_{\gamma_1})) \le 0 \quad \text{(A2)}$$

If $\gamma_1 = 1$, we can obtain the result directly from (A1), otherwise, we can divide (A1) by $1-\gamma_1$ and divide (A2) by $1-\gamma_2$ add them together to obtain

$$(SYY(\beta_{\gamma_1}) - SYY(\beta_{\gamma_2}))(\frac{\gamma_1}{1-\gamma_1} - \frac{\gamma_2}{1-\gamma_2}) \leq 0$$

Since $\gamma_1 > \gamma_2$, then $\frac{\gamma_1}{1-\gamma_1} > \frac{\gamma_2}{1-\gamma_2}$, thus $S_{YY}(\beta_{\gamma_1}) \leq S_{YY}(\beta_{\gamma_2})$.

Hence we have proven that $S_{YY}(\beta_\gamma)$ is a continuous monotonic function of γ.

In a constrained model, $\forall c$, there exists an estimator $\hat{\beta}_c$ that will minimize $\sum_{i=1}^{n}(X_i - \tilde{X}_i)^2$. Let $c^* = S_{YY}(\hat{\beta}_c)$, then $c^* \leq c$. Furthermore $S_{XX}(\hat{\beta}_c) \leq S_{XX}(\hat{\beta})$, $\forall \hat{\beta}$ such that $S_{YY}(\hat{\beta}) \leq c$ (A 3)

If $c > SYY(\beta_{\gamma=1})$, then the constraint can be ignored, and the estimator is the one that will minimize $\sum_{i=1}^{n}(X_i - \tilde{X}_i)^2$. If $c < S_{YY}(\beta_{\gamma=0})$, then no estimator would satisfy the constrained function. Hence we obtain $S_{YY}(\beta_{\gamma=0}) \leq c \leq S_{YY}(\beta_{\gamma=1})$.

The conclusion follows immediately if $c = S_{YY}(\beta_{\gamma=1})$ or $c = S_{YY}(\beta_{\gamma=0})$.

Hence, we only consider $S_{YY}(\beta_{\gamma=0}) < c < S_{YY}(\beta_{\gamma=1})$. Since $S_{YY}(\beta_\gamma)$ is a continuous and monotonic function of γ and according to the mean value theorem, there exists $0 < \gamma^* < 1$, such that $S_{YY}(\hat{\beta}_{\gamma^*}) = S_{YY}(\hat{\beta}_c)$ .Since $\hat{\beta}_{\gamma^*}$ minimizes the compound function when $\gamma = \gamma^*$, we have: $(1-\gamma^*)S_{XX}(\hat{\beta}_{\gamma^*}) + \gamma^* S_{YY}(\hat{\beta}_{\gamma^*}) \leq (1-\gamma^*)S_{XX}(\hat{\beta}_c) + \gamma^* S_{YY}(\hat{\beta}_c)$

In addition we have $S_{YY}(\hat{\beta}_{\gamma^*}) = S_{YY}(\hat{\beta}_c)$, and thus $S_{XX}(\hat{\beta}_{\gamma^*}) \leq S_{XX}(\hat{\beta}_c)$. We also have: $S_{XX}(\hat{\beta}_c) \leq S_{XX}(\hat{\beta})$, $\forall \hat{\beta}$ such that $S_{YY}(\hat{\beta}) \leq c$ from (A3).

Hence $S_{XX}(\hat{\beta}_{\gamma^*}) = S_{XX}(\hat{\beta}_c)$. The theorem is proven.

3) Now we derive the corresponding γ in the compound regression given a particular

value for c in the constrained regression.

$$\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2=\sum_{i=1}^{n}(Y_i-\bar{Y}+\hat{\beta}_1(\bar{X}-X_i))^2$$

$$=\sum_{i=1}^{n}(Y_i-\bar{Y})^2+\hat{\beta}_1^2\sum_{i=1}^{n}(\bar{X}-X_i)^2+2\hat{\beta}_1\sum_{i=1}^{n}(Y_i-\bar{Y})(\bar{X}-X_i)=S_{YY}+\hat{\beta}_1^2S_{XX}-2\hat{\beta}_1S_{XY}$$

$$\Rightarrow\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2=S_{YY}+\hat{\beta}_1^2S_{XX}-2\hat{\beta}_1S_{XY}\le c$$

$$\Rightarrow\hat{\beta}_1\in[\frac{S_{XY}-\sqrt{S_{XY}^2-S_{XX}(S_{YY}-c)}}{S_{XX}},\frac{S_{XY}+\sqrt{S_{XY}^2-S_{XX}(S_{YY}-c)}}{S_{XX}}]$$

From Theorem 1, we can see that:

If $\mathrm{S_{XY}}\ge 0$, then γ would be a decreasing function of $\hat{\beta}_1$; if $S_{XY}<0$, then γ would be an increasing function of $\hat{\beta}_1$. Let $\gamma_1>\gamma_2$, from (A2) we have

$$(1-\gamma_2)(S_{XX}(\beta_{\gamma_2})-S_{XX}(\beta_{\gamma_1}))+\gamma_2(S_{YY}(\beta_{\gamma_2})-S_{YY}(\beta_{\gamma_1}))\le 0$$

And we know $S_{YY}(\beta_{\gamma_2})\ge S_{YY}(\beta_{\gamma_1})$ from the proof of above theorem, hence the second part is non-negative; and hence the first part should be non-positive. Therefore, we have $S_{XX}(\beta_{\gamma_2})\le S_{XX}(\beta_{\gamma_1})$. That is, $\sum_{i=1}^{n}(X_i-\hat{X}_i)^2$ increases when γ increases, we should choose γ as small as possible. Therefore, we obtain:

$$\hat{\gamma}=\frac{S_{YY}-\hat{\beta}_1S_{XY}}{S_{YY}-\hat{\beta}_1S_{XY}+\hat{\beta}_1^4S_{XX}-\hat{\beta}_1^3S_{XY}},\ \hat{\beta}_1=\frac{S_{XY}+sign(S_{XY})\sqrt{S_{XY}^2-S_{XX}(S_{YY}-c)}}{S_{XX}}$$

Theorem 4. (a) The Geometric Mean Regression would always yield equal efficiencies for the estimations of X and Y respectively. (b) The Ordinary Least Squares Regressions for X and Y have the same efficiencies, albeit in reverse order, for X and Y.

Proof. (a) As described above $\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 = S_{YY}+\hat{\beta}_1^2 S_{XX}-2\hat{\beta}_1 S_{XY}$

$$\sum_{i=1}^{n}(X_i-\hat{X}_i)^2 = \frac{1}{\hat{\beta}_1^2}\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 = \frac{1}{\hat{\beta}_1^2}S_{YY}+S_{XX}-2\frac{1}{\hat{\beta}_1}S_{XY}$$

For geometric mean regression, we have $\hat{\beta}_1 = sign(S_{XY})\sqrt{S_{YY}/S_{XX}}$; hence,

$$e_1 = \frac{\min\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2} = \frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 \,|_{\hat{\beta}_1=\frac{S_{XY}}{S_{XX}}}}{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 \,|_{\hat{\beta}_1=\sqrt{\frac{S_{YY}}{S_{XX}}}}} = \frac{S_{YY}+\frac{S_{XY}^2}{S_{XX}}-2\frac{S_{XY}^2}{S_{XX}}}{S_{YY}+S_{YY}-2S_{XY}\sqrt{\frac{S_{YY}}{S_{XX}}}} = \frac{S_{XX}S_{YY}-S_{XY}^2}{2S_{XX}S_{YY}-2S_{XY}\sqrt{S_{XX}S_{YY}}}$$

$$e_2 = \frac{\min\sum_{i=1}^{n}(X_i-\hat{X}_i)^2}{\sum_{i=1}^{n}(X_i-\hat{X}_i)^2} = \frac{\frac{1}{\hat{\beta}_1^2}\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 \,|_{\hat{\beta}_1=\frac{S_{YY}}{S_{XY}}}}{\frac{1}{\hat{\beta}_1^2}\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 \,|_{\hat{\beta}_1=\sqrt{\frac{S_{YY}}{S_{XX}}}}} = \frac{S_{XX}S_{YY}-S_{XY}^2}{2S_{XX}S_{YY}-2S_{XY}\sqrt{S_{XX}S_{YY}}}$$

Thus we have proven that e$_1$=e$_2$, for the geometric mean regression.

(b) The equality of e$_1$ (for r = 0) and e$_2$ (for r = 1) are easily proven as follows:

$$\gamma=0, \beta_1=\frac{S_{YY}}{S_{XY}}, \quad e_1=\frac{S_{YY}+S_{XY}^2/S_{XX}-2S_{XY}^2/S_{XX}}{S_{YY}+S_{XX}\,S_{YY}^2/S_{XY}^2-2S_{YY}} = \frac{S_{YY}-S_{XY}^2/S_{XX}}{S_{YY}^2S_{XX}/S_{XY}^2-S_{YY}} = \frac{S_{YY}S_{XX}-S_{XY}^2}{S_{YY}^2S_{XX}^2/S_{XY}^2-S_{YY}S_{XX}}$$

$$\gamma=1, \beta_1=\frac{S_{XY}}{S_{XX}}, e_2=\frac{S_{XY}^2/S_{YY}+S_{XX}-2S_{XY}^2/S_{YY}}{S_{YY}S_{XX}^2/S_{XY}^2+S_{XX}-2S_{XX}} = \frac{S_{XX}-S_{XY}^2/S_{YY}}{S_{YY}S_{XX}^2/S_{XY}^2-S_{XX}} = \frac{S_{YY}S_{XX}-S_{XY}^2}{S_{XX}^2S_{YY}^2/S_{XY}^2-S_{YY}S_{XX}}$$

Appendix B: Programs

1. MatLab code for finding slope estimates of constrained regression simulation

```
c=0;
for i=1:101;
    for j=1:101;
        if e1(i,j)>=0.7 & e2(i,j)>=0.5 & e3(i,j)>=c;
            c=e3(i,j);
            k=i;
            l=j;
        else;
        end;
    end;
end;
```

2. MatLab code for resampling simulation

```
function [up,m,low]=resamp2d(what);
A=what(:,1);
B=what(:,2);
for j=0:1:10;
   gamma=j/10;
for i=1:1000;
   s=rand(1,200)*200+1;
   k=fix(s);
   x=what(k,1);
   y=what(k,2);
f=inline('(gamma+(1-gamma)/beta^2)*(s(y,y)+beta^2*s(x,x)-2*beta*s(x,y))
','beta','gamma','x','y');
beta(j+1,i)=fminsearch(f,1,[],gamma,x,y);
end;
m(j+1)=mean(beta(j+1,:));
st(j+1)=std(beta(j+1,:));
up(j+1)=m(j+1)+1.96*st(j+1);
low(j+1)=m(j+1)-1.96*st(j+1);
end;
```

3. MatLab code for simulation data generation

```
function pt=sim3d;
p = 200;    % Designed sample size
NoiseLevelX=2*3;
NoiseLevelY=1.41*3;
NoiseLevelZ=1*3;
trueSlopeX=2;
trueSlopeY=3;
dataRange=3;
pt=getNoise3D(p,NoiseLevelX,NoiseLevelY,NoiseLevelZ,...
   trueSlopeX,trueSlopeY,dataRange);


function pt=myData(p,NoiseLevelX,NoiseLevelY,NoiseLevelZ,...
   trueSlopeX,trueSlopeY,dataRange)
pt=zeros(p,3);
sigmaX = NoiseLevelX*rand(p,1);
sigmaY = NoiseLevelY*rand(p,1);
sigmaZ = NoiseLevelZ*rand(p,1);
pt(:,1) = dataRange*randn(p,1);
pt(:,2) = dataRange*randn(p,1);
pt(:,3) = trueSlopeX*pt(:,1)+trueSlopeY*pt(:,2)+sigmaZ;
pt(:,1) = pt(:,1)+sigmaX;
pt(:,2) = pt(:,2)+sigmaY;
return;
```