On Seemingly Unrelated Semiparametric Models

Roozbeh, Mahdi
Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad
Mashhad, P.O. Box 91775-1159,Iran
E-mail: m.roozbeh.stat@gmail.com

Arashi, Mohammad
Faculty of Mathematics, Shahrood University of Technology
Shahrood, P.O. Box 316-3619995161, Iran
E-mail: m.arash.stat@yahoo.com

Niroumand, Hosseinali
Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad
Mashhad, P.O. Box 91775-1159,Iran
E-mail: nirumand@math.um.ac.ir

1 Introduction

A seemingly unrelated regression (SUR) system proposed by Zellner (1962) comprises several individual relationships that are linked by the fact that their disturbances are correlated. Such models have found many applications. For example, demand functions can be estimated for different households (or household types) for a given commodity. The correlation among the equation disturbances could come from several sources such as correlated shocks to household income. Alternatively, one could model the demand of a household for different commodities, but adding-up constraints leads to restrictions on the parameters of different equations in this case. On the other hand, equations explaining some phenomenon in different cities, states, countries, firms or industries provide a natural application as these various entities are likely to be subject to spill overs from economy-wide or world-wide shocks. In fact the SUR model is a generalization of multivariate regression using a vectorized parameter model. There are two main motivations for use of SUR. The first one is to gain efficiency in estimation by combining information on different equations. The second motivation is to impose and/or test restrictions that involve parameters in different equations.

In most of the empirical works people are often concerned about problems with the specification of the model or problems with the data. This problem arises in situations when the explanatory variables are highly inter-correlated.

In this paper, we apply differencing method to remove the nonparametric parts in seemingly unrelated semiparametric (SUS) model and then estimate the linear parts to accelerate estimating the parametric parts (linear parts). At the second step, nonparametric technique is applied to estimate the nonparametric parts. In this regard we deal with SUS model applying differencing methodology, under multicollinearity setting. Thus we use ridge regression concept that was proposed in the 1970's to combat the multicollinearity in regression problems.

2 Specifying SUS Model

Consider a system of M equations that follow the semiparametric model given by

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{f}_i(\boldsymbol{z}) + \boldsymbol{\epsilon}_i, \quad i = 1, ..., M \tag{2.1}$$

where $\boldsymbol{Y}_i$ is a vector of $T \times 1$ observations on the dependent variable, $\boldsymbol{X}_i$ is a $T \times p_i$ matrix of observations on p_i explanatory variables, $\boldsymbol{\beta}_i$ is a $p_i \times 1$ dimensional vector of unknown location parameters, $\boldsymbol{f}_i(\boldsymbol{z})$ is a $T \times 1$ vector and $\boldsymbol{\epsilon}_i$ is a $T \times 1$ vector of random errors. M is the number of equations in the system, T is the number of observations per equation and p_i is the number of rows in the vector $\boldsymbol{\beta}_i$.

The M equations in (2.1) can be written in the compact form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{f}(\boldsymbol{z}) + \boldsymbol{\epsilon}, \tag{2.2}$$

where $\boldsymbol{Y} = (\boldsymbol{Y}'_1, \boldsymbol{Y}'_2, ..., \boldsymbol{Y}'_M)'$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}'_1, \boldsymbol{\epsilon}'_2, ..., \boldsymbol{\epsilon}'_M)'$ and $\boldsymbol{f}(\boldsymbol{z}) = (\boldsymbol{f}'_1(\boldsymbol{z}), \boldsymbol{f}'_2(\boldsymbol{z}), ..., \boldsymbol{f}'_M(\boldsymbol{z}))'$ are each of dimension $TM \times 1$, $\boldsymbol{X} = diag(\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_M)$ is of dimension $TM \times p$, $p = \sum_{i=1}^{M} p_i$, and $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, ..., \boldsymbol{\beta}'_M)'$ is a vector of $p \times 1$ parameters.

Furthermore, throughout we will have the following assumptions:

Assumption 1. $\boldsymbol{X_i}$ is fixed with rank p_i,

Assumption 2. plim $\frac{1}{T}(\boldsymbol{X'_i X_i})$ is non-singular with finite and fixed elements.

Assumption 3. plim $\frac{1}{T}(\boldsymbol{X'_i X_j})$ is non-singular with finite and fixed elements.

Assumption 4. $E(\epsilon_i \epsilon'_j) = v_{ij}\boldsymbol{I}_T$, where v_{ij} designate the covariance between the i^{th} and j^{th} equations for each observation in the sample. The above expression can be also written as

$$E(\boldsymbol{\epsilon}) = \boldsymbol{0}, \; E(\boldsymbol{\epsilon\epsilon'}) = \boldsymbol{V}, \tag{2.3}$$

$$\boldsymbol{V} = \begin{pmatrix} v_{11} & v_{12} & ... & v_{1M} \\ v_{21} & v_{22} & ... & v_{2M} \\ \vdots & \ddots & & \vdots \\ v_{M1} & v_{M2} & ... & v_{MM} \end{pmatrix} \bigotimes \boldsymbol{I}_T = \begin{pmatrix} \boldsymbol{V}_{11} = v_{11}\boldsymbol{I}_T & \boldsymbol{V}_{12} = v_{12}\boldsymbol{I}_T & ... & \boldsymbol{V}_{1M} = v_{1M}\boldsymbol{I}_T \\ \boldsymbol{V}_{21} = v_{21}\boldsymbol{I}_T & \boldsymbol{V}_{22} = v_{22}\boldsymbol{I}_T & ... & \boldsymbol{V}_{2M} = v_{2M}\boldsymbol{I}_T \\ \vdots & \ddots & & \vdots \\ \boldsymbol{V}_{M1} = v_{M1}\boldsymbol{I}_T & \boldsymbol{V}_{M2} = v_{M2}\boldsymbol{I}_T & ... & \boldsymbol{V}_{MM} = v_{MM}\boldsymbol{I}_T \end{pmatrix},$$

is a $TM \times TM$ positive definite symmetric matrix and $\otimes$ represents the kronecker product. Thus the errors of each equation are assumed to be homoscedastic and not autocorrelated, but it is also assumed that there is contemporaneous correlation between corresponding errors in different equations.

Semiparametric models are more flexible than standard linear models since they have a parametric and a nonparametric component. They can be a suitable choice when one suspects that the response y linearly depends on x, but that it is nonlinearly related to z. Yatchew (1997) concentrates on estimation of the linear component of a semiparametric model and used differencing to eliminate bias induced from the presence of the nonparametric component. Wang et al. (2007) used higher order differences for optimal efficiency in estimating the linear part by using a special class of difference sequences. In continuation, let $\boldsymbol{d} = (d_0, ..., d_m)$ be a $m+1$ vector, where m is the order of differencing and $d_0, ..., d_m$ are differencing weights satisfying the conditions

$$\sum_{j=0}^{m} d_j = 0, \quad \sum_{j=0}^{m} d_j^2 = 1. \tag{2.4}$$

Now, we define the $(T-m) \times T$ differencing matrix $\boldsymbol{D}$ whose elements satisfy (2.4) as (see Yatchew, 2003 for some examples)

$$\boldsymbol{D} = \begin{pmatrix} d_0 & d_1 & ... & d_m & 0 & 0 & ... & 0 \\ 0 & d_0 & d_1 & ... & d_m & 0 & ... & 0 \\ \vdots & \ddots & & & & & & \vdots \\ 0 & 0 & ... & 0 & d_0 & d_1 & ... & d_m \end{pmatrix}.$$

Imposing the differencing matrix to the model (2.2), permits direct estimation of the parametric effect. In particular, it takes

$$(\boldsymbol{I}_M \otimes \boldsymbol{D})\boldsymbol{Y} = (\boldsymbol{I}_M \otimes \boldsymbol{D})\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_M \otimes \boldsymbol{D})\boldsymbol{f}(\boldsymbol{z}) + (\boldsymbol{I}_M \otimes D)\boldsymbol{\epsilon}. \tag{2.5}$$

Since the data have been reordered so that the z's are close, the application of the differencing matrix $\boldsymbol{D}$ in model (2.5) removes the nonparametric effect in large samples (Yatchew, 2000). Thus, the underlying model is rewritten as

$$\widetilde{\boldsymbol{Y}} \doteq \widetilde{\boldsymbol{X}}\boldsymbol{\beta} + \widetilde{\boldsymbol{\epsilon}}, \tag{2.6}$$

where $\widetilde{\boldsymbol{Y}} = (\boldsymbol{I}_M \otimes \boldsymbol{D})\boldsymbol{Y}$, $\widetilde{\boldsymbol{X}} = (\boldsymbol{I}_M \otimes \boldsymbol{D})\boldsymbol{X}$, $\widetilde{\boldsymbol{\epsilon}} = (\boldsymbol{I}_M \otimes \boldsymbol{D})\boldsymbol{\epsilon}$.

3 Difference-Based Ridge Estimator

It is well-known that adopting the linear model (2.6), the unbiased estimator of $\boldsymbol{\beta}$ is the following generalized difference-based estimator for SUS model given by

$$\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GD}} = \boldsymbol{C_D}^{-1}\widetilde{\boldsymbol{X}}'\boldsymbol{V_D}^{-1}\widetilde{\boldsymbol{Y}},\ \boldsymbol{C_D} = \widetilde{\boldsymbol{X}}'\boldsymbol{V_D}^{-1}\widetilde{\boldsymbol{X}}. \tag{3.1}$$

It is observed from (3.1) that the properties of the generalized difference-based estimator of $\boldsymbol{\beta}$ depend heavily on the characteristics of the information matrix $\boldsymbol{C_D}$. If the $\boldsymbol{C_D}$ matrix is ill-conditioned (near dependency among various columns of $\boldsymbol{C_D}$), then the $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GD}}$ produce unduly large sampling variances. Moreover, some of the regression coefficients may be statistically insignificant with wrong sign and meaningful statistical inference become difficult for the researcher. As a remedy, following Hoerl and Kennard (1970), we suggest to use the following estimator namely generalized difference-based ridge estimator in SUS model

$$\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GD}}(\boldsymbol{K}) = \boldsymbol{C_D}(\boldsymbol{K})^{-1}\widetilde{\boldsymbol{X}}'\boldsymbol{V_D}^{-1}\widetilde{\boldsymbol{Y}}, \boldsymbol{C_D}(\boldsymbol{K}) = \boldsymbol{C_D} + \boldsymbol{K}, \boldsymbol{K} = diag(\boldsymbol{K}_1, \boldsymbol{K}_2, ..., \boldsymbol{K}_M), \tag{3.2}$$

where $\boldsymbol{K}_i = k_i\boldsymbol{I}_{p_i}, k_i \geq 0$ are the shrinking parameter for $i = 1, .., M$.

Now consider the following set of linear non-stochastic constraints

$$\boldsymbol{R}_i\boldsymbol{\beta}_i = \boldsymbol{r}_i, \quad i = 1, ..., M \tag{3.3}$$

for a given $q_i \times p_i$ matrix $\boldsymbol{R}_i$ with rank $q_i < p_i$ and a given $q_i \times 1$ vector $\boldsymbol{r}_i$. The full row rank assumptions are chosen for convenience and can be justified by the fact that every consistent linear equation can be transformed into an equivalent equation with a coefficient matrix of full row rank. Subject to the linear restrictions (3.3), the generalized restricted difference-based estimator is given by

$$\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}} = \hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GD}} - \boldsymbol{C_D}^{-1}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{C_D}^{-1}\boldsymbol{R}')^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GD}} - \boldsymbol{r}), \tag{3.4}$$

where $\boldsymbol{R} = diag(\boldsymbol{R}_1, \boldsymbol{R}_2, ..., \boldsymbol{R}_M)$ and $\boldsymbol{r} = (\boldsymbol{r}'_1, \boldsymbol{r}'_2, ..., \boldsymbol{r}'_M)'$ are $q \times p$ and $q \times 1$ matrices ($q = \sum_{i=1}^{M} q_i$).

So, the generalized restricted difference-based ridge estimator in SUS model can be written as

$$\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K}) = \hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GD}}(\boldsymbol{K}) - \boldsymbol{C_D}(\boldsymbol{K})^{-1}\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{C_D}(\boldsymbol{K})^{-1}\boldsymbol{R}']^{-1}[\boldsymbol{R}\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GD}}(\boldsymbol{K}) - \boldsymbol{r}]. \tag{3.5}$$

Then it is easy to see that the proposed estimators are restricted with respect to $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$. It is also clear that for $\boldsymbol{K} = \boldsymbol{0}$, we get $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{0}) = \hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}$.

4 Biases and MSE Expressions

In this section, we calculate the risk function for the proposed estimator given in previous section. For calculating the risk function of the proposed estimators in the last Section and driving a necessary and sufficient condition for the superiority of $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{FGRD}}(\boldsymbol{K})$ over $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{FGD}}$, first we obtain a new formula for $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K})$ which simplifies the calculation of risk function as follows:

$$\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K}) = \boldsymbol{N_D}(\boldsymbol{K})\widetilde{\boldsymbol{X}}'\boldsymbol{V}_{\boldsymbol{D}}^{-1}\widetilde{\boldsymbol{Y}} - \boldsymbol{N_D}(\boldsymbol{K})\boldsymbol{C_D}(\boldsymbol{K})\boldsymbol{\beta}_0 + \boldsymbol{\beta}_0, \tag{4.1}$$

where, $\boldsymbol{\beta}_0 = \boldsymbol{R}'(\boldsymbol{R}\boldsymbol{R}')^{-1}\boldsymbol{r}$ and

$$\boldsymbol{N_D}(\boldsymbol{K}) = \boldsymbol{C_D}(\boldsymbol{K})^{-1} - \boldsymbol{C_D}(\boldsymbol{K})^{-1}\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{C_D}(\boldsymbol{K})^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{C_D}(\boldsymbol{K})^{-1}. \tag{4.2}$$

Then, we can calculate the bias and covariance matrix of $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K})$ by using equation (4.1) as follows:

$$E[\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K}) - \boldsymbol{\beta}] = -\boldsymbol{K}\boldsymbol{N_D}(\boldsymbol{K})\boldsymbol{\beta}, \tag{4.3}$$

and

$$Cov[\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K})] = \boldsymbol{N_D}(\boldsymbol{K})\boldsymbol{C_D}\boldsymbol{N_D}(\boldsymbol{K}). \tag{4.4}$$

With direct calculation using $\boldsymbol{C_D} = \boldsymbol{C_D}(\boldsymbol{K}) - \boldsymbol{K}$, we conclude that $\boldsymbol{N_D}(\boldsymbol{K})\boldsymbol{C_D}\boldsymbol{N_D}(\boldsymbol{K}) = \boldsymbol{N_D}(\boldsymbol{K}) - \boldsymbol{N_D}(\boldsymbol{K})\boldsymbol{K}\boldsymbol{N_D}(\boldsymbol{K})$ and in particular, $\boldsymbol{N_D}(\boldsymbol{0})\boldsymbol{C_D}\boldsymbol{N_D}(\boldsymbol{0}) = \boldsymbol{N_D}(\boldsymbol{0})$ when $\boldsymbol{K} = \boldsymbol{0}$. Therefore, the risk function of $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K})$ and $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}$ are

$$MSE[\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K}), \boldsymbol{\beta}] = \boldsymbol{N_D}(\boldsymbol{K}) - \boldsymbol{N_D}(\boldsymbol{K})\boldsymbol{K}\boldsymbol{N_D}(\boldsymbol{K}) + \boldsymbol{K}\boldsymbol{N_D}(\boldsymbol{K})\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{N_D}(\boldsymbol{K})\boldsymbol{K}, \tag{4.5}$$

$$MSE[\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}, \boldsymbol{\beta}] = \boldsymbol{N_D}(\boldsymbol{0}). \tag{4.6}$$

5 MSE-Superiority

In this section, we provide necessary and sufficient conditions for which the difference-based ridge estimator $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K})$ out performs the differencing estimator $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}$ in the sense of having smaller MSE.

From (4.5) and (4.6), under $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$ the MSE difference is given by

$$\begin{aligned} \boldsymbol{\Delta} &= MSE(\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}, \boldsymbol{\beta}) - MSE[\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K}), \boldsymbol{\beta}] \\ &= \boldsymbol{N_D}(\boldsymbol{0}) - \boldsymbol{N_D}(\boldsymbol{K}) + \boldsymbol{N_D}(\boldsymbol{K})\boldsymbol{K}\boldsymbol{N_D}(\boldsymbol{K}) - \boldsymbol{K}\boldsymbol{N_D}(\boldsymbol{K})\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{N_D}(\boldsymbol{K})\boldsymbol{K}. \end{aligned} \tag{5.1}$$

Lemma 5.1. *The $Cov[\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K})]$ is consistently smaller than $Cov(\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}})$, that is to say the following inequality always holds for any arbitrary $\boldsymbol{K} > \boldsymbol{0}$*

$$\tilde{\boldsymbol{\Delta}} = Cov(\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}) - Cov[\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K})] > 0.$$

Moreover, $\tilde{\boldsymbol{\Delta}}$ is monotonously increased with respect to $\boldsymbol{K}$.

Theorem 5.1. *Let us be given the estimator $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K})$ under the linear regression model with true restrictions $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$. The MSE difference $\boldsymbol{\Delta}$ is nonnegative definite if and only if*

$$\boldsymbol{\beta}'\boldsymbol{G}^{+}\boldsymbol{\beta} \leq 1, \tag{5.2}$$

where $\boldsymbol{G} = 2k^{-1}\boldsymbol{P} + (\boldsymbol{P}\boldsymbol{C_D}\boldsymbol{P})^{+}$.

It is important to note that without taking the restriction $k_1 = k_2 = \cdots = k_M = k > 0$, obtaining the MSE superiority is not an easy job. However, one can extend this result to a more general case for further research.

6 Upper Bound For Ridge Parameter

In the process of determining $\boldsymbol{K}$, on one side, we must control the condition number of $\boldsymbol{C_D}(\boldsymbol{K})$ to a lesser level if we want to avoid the instability of estimated coefficients brought by the morbidity of $\boldsymbol{C_D}$. Hence, we must do our best to let the ridge matrix $\boldsymbol{K}$ be big. Furthermore, we know the bigger the $\boldsymbol{K}$ is, the smaller the covariance of the estimated coefficients is. It implies that the estimator to be more stable. On the other side, we know, in view of the biased estimator, when the $\boldsymbol{K}$ is smaller, the estimator will be better (the $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K})$ will be apart badly from the actual $\boldsymbol{\beta}$ as $\boldsymbol{K}$ increasing). In other words, we must comply some principles to select $\boldsymbol{K}$.

Belsley et al. (1980) proposed that the multicollinearity would take effect apparently as the condition number of $\boldsymbol{C_D}$ is bigger than 10. The correlation of the variables of $\widetilde{\boldsymbol{X}}$ is strong comparatively when the condition number of $\boldsymbol{C_D}$ is between 30 and 100. While the condition number is bigger than 100, the correlation would become very strong and the estimated coefficients is very unstable. They suggested that we can obtain the good and stable coefficients if the condition number of $\boldsymbol{C_D}$ is not bigger than 10. Based on this criterion, we can easily choose the k such that the condition number of $\boldsymbol{C_D}$ is reduced to 10 (see Liu, 2003).

Although the criterion mentioned above is simple, our problem to select $\boldsymbol{K}$ is not yet completely solved. Therefore, we give a range to select $\boldsymbol{K}$ in Theorem 5.1.

Theorem 6.1. *Let us be given the estimator $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K})$ under the linear regression model with true restrictions $\boldsymbol{R\beta} = \boldsymbol{r}$ and $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. The MSE difference $\boldsymbol{\Delta}$ is nonnegative definite if*

$$k \leq \frac{2}{\boldsymbol{\beta}'\boldsymbol{P}\boldsymbol{\beta}}. \tag{6.1}$$

7 Monte Carlo Study

In this section, we proceed to comparison of the proposed estimators numerically. We will compare the trace of $MSE[\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K}), \boldsymbol{\beta}]$ and $MSE(\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}, \boldsymbol{\beta})$ and define scalar Δ as

$$\Delta = tr(\boldsymbol{\Delta}) = tr[\boldsymbol{N_D}(\boldsymbol{0})] - tr[\boldsymbol{N_D}(\boldsymbol{K})] + tr[\boldsymbol{N_D}(\boldsymbol{K})\boldsymbol{K}\boldsymbol{N_D}(\boldsymbol{K})] - \boldsymbol{\beta}'\boldsymbol{N_D}(k)\boldsymbol{K}^2\boldsymbol{N_D}(\boldsymbol{K})\boldsymbol{\beta}. \tag{7.1}$$

Our sampling experiment consists of different combinations of k_i. In this study we simulate the response for $M = 3$ equations and $T = 100$ with 10^5 iterations from the following model:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{f(z)} + \boldsymbol{\epsilon}, \tag{7.2}$$

where $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_3 = (1.5, 2, 3, -1, 5)$, for $i = 1, ..., T$, $\boldsymbol{x}_i \sim N_5(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ with

$$\boldsymbol{\mu_x} = \begin{pmatrix} 2.5 \\ 2 \\ 3 \\ 1 \\ -1 \end{pmatrix}, \quad \boldsymbol{\Sigma_x} = \begin{pmatrix} 1 & r_x & r_x & r_x & r_x \\ r_x & 1 & r_x & r_x & r_x \\ r_x & r_x & 1 & r_x & r_x \\ r_x & r_x & r_x & 1 & r_x \\ r_x & r_x & r_x & r_x & 1 \end{pmatrix},$$

where $r_x = 0.99$, $\boldsymbol{\epsilon} \sim N_{TM}(\boldsymbol{0}, \boldsymbol{V})$ which $v_{ii} = 1$ and $v_{ij} = r_\epsilon = 0.95$ for $i, j = 1, ..., M$, $i \neq j$. We consider the following three functions as nonparametric parts:

$$f_1(z_i) = \sqrt{z_i(1-z_i)}\sin\left(\frac{2.1\pi}{z_i + 0.05}\right), f_2(z_i) = \frac{1}{9}\sum_{j=1}^{9}\phi(z_i; j, (j/10)^j), f_3(z_i) = \frac{1}{6}\sum_{j=1, j\neq 4,6}^{8}\phi\Big(z_i; j, \big((j+2)/10\big)^j\Big),$$

where $\phi(x; \mu, \sigma^2)$ is a normal density function with mean μ and variance σ^2 for $z_i = 10i/T$.

The main reason for taking such functions is to determine how the estimators are optimal in the sense of making better coverage. In model (2.6) the parametric effect, $\boldsymbol{\beta}$, is estimated by a differencing

Table 1: Matrices of linear restrictions used during simulation study

R_1					R_2					R_3				
1	2	4	5	2	-2	1	4	5	5	1	1	4	5	5
-2	-1	3	1	2	-1	2	3	-3	0	-1	-1	3	3	1
-1	2	3	1	2	1	2	4	-1	-2	-2	3	8	0	1
1	3	2.1	-1	4.13	3	2	5	-1	4.75	1	1	3	-2	3

Table 2: Evaluation of estimators at different values k_i for model (7.2)

k_i Coefficients	0	1	2	3	4	5	6	7	8
$Eq.1$									
$\hat{\beta}_1$	1.54858	1.54433	1.54019	1.53616	1.53223	1.52840	1.52467	1.52103	1.51748
$\hat{\beta}_2$	1.98380	1.98522	1.98660	1.98794	1.98925	1.99053	1.99177	1.99298	1.99417
$\hat{\beta}_3$	3.07231	3.06599	3.05982	3.05382	3.04798	3.04228	3.03672	3.03130	3.02602
$\hat{\beta}_4$	-1.04237	-1.03866	-1.03505	-1.03153	-1.02811	-1.02477	-1.02151	-1.01834	-1.01524
$\hat{\beta}_5$	4.95319	4.95729	4.96127	4.96516	4.96894	4.97263	4.97622	4.97973	4.98315
$Eq.2$									
$\hat{\beta}_1$	1.49880	1.49839	1.49799	1.49759	1.49720	1.49681	1.49643	1.49605	1.49568
$\hat{\beta}_2$	1.98797	1.98388	1.97983	1.97584	1.97190	1.96802	1.96418	1.96039	1.95665
$\hat{\beta}_3$	3.00556	3.00745	3.00932	3.01117	3.01299	3.01479	3.01656	3.01832	3.02005
$\hat{\beta}_4$	-1.00205	-1.00275	-1.00344	-1.00412	-1.00479	-1.00546	-1.00611	-1.00676	-1.00740
$\hat{\beta}_5$	4.99953	4.99937	4.99921	4.99905	4.99890	4.99875	4.99860	4.99845	4.99830
$Eq.3$									
$\hat{\beta}_1$	1.80621	1.80637	1.80653	1.80669	1.80686	1.80702	1.80718	1.80734	1.80750
$\hat{\beta}_2$	1.96062	1.96060	1.96058	1.96056	1.96054	1.96052	1.96050	1.96048	1.96046
$\hat{\beta}_3$	3.11373	3.11379	3.11385	3.11391	3.11397	3.11403	3.11409	3.11415	3.11421
$\hat{\beta}_4$	-0.96500	-0.96498	-0.96496	-0.96494	-0.96493	-0.96491	-0.96489	-0.96487	-0.96485
$\hat{\beta}_5$	4.82064	4.82055	4.82045	4.82036	4.82026	4.82017	4.82007	4.81998	4.81989
$b'(\hat{\boldsymbol{\beta}})b(\hat{\boldsymbol{\beta}})$	0	8.742e-05	3.417e-04	7.515e-04	1.306e-03	1.996e-03	2.812e-03	3.746e-03	4.790e-03
$tr[Cov(\hat{\boldsymbol{\beta}})]$	0.021229	0.020779	0.020346	0.019927	0.019523	0.019132	0.018754	0.018389	0.018036
$tr[MSE(\hat{\boldsymbol{\beta}},\boldsymbol{\beta})]$	0.021229	0.020867	0.020687	0.020678	0.020829	0.021128	0.021567	0.022136	0.022826
Δ	0	0.000362	0.000541	0.000550	0.000399	0.000100	-0.000338	-0.000906	-0.001597
$mse[\hat{\boldsymbol{f}}(\boldsymbol{z}),\boldsymbol{f}(\boldsymbol{z})]$	0.405953	0.402391	0.399008	0.395794	0.392739	0.389835	0.387073	0.384446	0.381905

procedure. Optimal differencing weights do not have analytic expressions but may be calculated easily using an optimization routine. Hall et al. (1990) present weights to order $m = 10$. These contain some minor errors. We use a fourth-order differencing coefficients, $d_0 = 0.8873$, $d_1 = -0.3099$, $d_2 = -0.2464$, $d_3 = -0.1901$, and $d_4 = -0.1409$ in which case $m = 4$.

All computations were conducted using the R statistical package. The matrix $\boldsymbol{C_D}$ has the smallest and largest eigenvalues respectively given by: $\lambda_{min} = 52.23$, $\lambda_{max} = 6914.11$. Thus the ratio is equal to $\lambda_{max}/\lambda_{min} = 132.37$, which shows an existence of multicollinearity in the data set.

After 10^5 samples are generated, the estimation of parameters, square of bias, trace of covariance, trace of MSE and mse of nonparametric estimation are computed for different values of $k_1 = k_2 = k_3 = 0, 1, ..., 8$ in model (7.2). The results are summarized in Tables 2 and 3.

In Table 3, the $\hat{\Delta}$ for different values of $k_i = 0, 1, ..., 7$, when $k_j = k_l = 0$, for considering the trace of each k_i on the $\hat{\Delta}$ is computed.

In Figure 1, the $\hat{\Delta}$ versus $k = k_1 = k_2 = k_3$ is plotted. In Figure 2, to ponder the behavior of each ridge parameter k_i, solely, the $\hat{\Delta}$ versus each ridge parameter k_i, when $k_j = k_l = 0$ is plotted. According to Figure 2, it can be realized that for all combinations of k_i, $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K})$ is better than

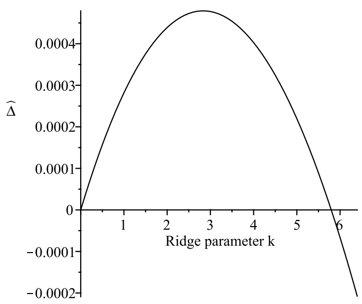

Figure 1: The diagram of $\hat{\Delta}$ versus k.

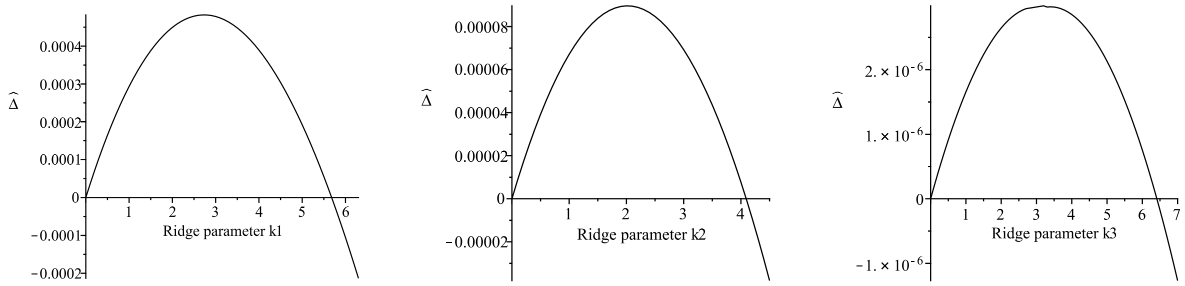

Figure 2: The diagram of $\hat{\Delta}$ versus k_i.

Table 3: $\hat{\Delta}$ values for different values of k_i

(k_1, k_2, k_3)	(0,0,0)	(1,0,0)	(2,0,0)	(3,0,0)	(4,0,0)	(5,0,0)	(6,0,0)	(7,0,0)
$\hat{\delta}$	0	2.9340e-04	4.4936e-04	4.7824e-04	3.8959e-04	1.9226e-04	-1.0554e-04	-4.9625e-04
(k_1, k_2, k_3)	(0,0,0)	(0,1,0)	(0,2,0)	(0,3,0)	(0,4,0)	(0,5,0)	(0,6,0)	(0,7,0)
$\hat{\Delta}$	0	6.7112e-05	8.9622e-05	6.9268e-05	7.7177e-06	-9.3424e-05	-2.3261e-04	-8.5577e-04
(k_1, k_2, k_3)	(0,0,0)	(0,0,1)	(0,0,2)	(0,0,3)	(0,0,4)	(0,0,5)	(0,0,6)	(0,0,7)
$\hat{\Delta}$	0	1.6561e-06	2.6460e-06	2.9739e-06	2.6435e-06	1.6589e-06	2.3905e-08	-2.2576e-06

$\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}$ if $k_i \leq a_i^*$, which $a_i^* = \frac{2}{\boldsymbol{\beta}_i' \boldsymbol{P}_i \boldsymbol{\beta}_i}$, when $k_j = k_l = 0$, and it is equal to 5.68, 4.06 and 6.42 in equations 1, 2 and 3, respectively. Furthermore, in each case, the maximum of $\hat{\Delta}$ is obtained when k_i equals to median range of $(0, a_i^*)$ i.e., $\frac{a_i^*}{2} = \frac{1}{\boldsymbol{\beta}_i' \boldsymbol{P}_i \boldsymbol{\beta}_i}$ (the optimum values of k_i for superiority of $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}(\boldsymbol{K})$ with respect to $\hat{\boldsymbol{\beta}}^{SUS}_{\boldsymbol{GRD}}$).

REFERENCES

Ahn, H. and Powell, J. (1993). Semiparametric estimation of censored selection models with a non-parametric selection mechanism. *J. Econometrics*, **58**, 3-29.

Fiebig, D. G., (2001). Seemingly Unrelated Regression. In Baltagi, B. eds, *A Companion to Theoretical Econometrics*, Backwell Publishers, 0101-121.

Grob, J., (2003). Restricted ridge estimation. *Statist. Prob. Lett.*, **65**, 57-64.

Hall, P. Kay, J.W. and Titterington, D.M., (1990). On estimation of noise variance in two-dimensional signal processing. *Advanced in appl. Prob.* , **23**, 476-495.

Hoerl, A.E. and Kennard, R.W., (1970). Ridge regression biased estimation for non-orthogonal problems. *Technometrics* , **12**, 69-82.

Yatchew, A., (2003). *Semiparametric regression for the applied econometrician.* Cambridge University press, Cambridge.

Zellner, A., (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Amer. Statist. Assoc.*, **57**, 348-368.

RÉSUMÉ (ABSTRACT) This article considers estimation in the seemingly unrelated semiparametric (SUS) models, when the explanatory variables are affected by multicollinearity. It is also suspected that some additional linear constraints may hold on the whole parameter space. In sequel we propose difference-based and difference-based ridge type estimators combining the restricted least squares method in the model under study. Necessary and sufficient conditions for the superiority of the ridge type estimator over the non-ridge type estimator for selecting the ridge parameter $\boldsymbol{K}$ are derived. Lastly, a Monte Carlo simulation study is conducted to estimate the parametric and non-parametric parts. In this regard, local linear regression method for estimating the non-parametric function are used.