STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. 11, March 2023, pp 258–275.
Published online in International Academic Press (www.IAPress.org)

Feasible Stein-Type and Preliminary Test Estimations in the System Regression Model

Mina Norouzirad[1,*] Mohammad Arashi[2], Filipe. J. Marques[1,3], Naushad A. Mamode Khan[4]

[1]Center for Mathematics and Applications (NovaMath), FCT NOVA, Portugal
[2]Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Iran
[3]Department of Mathematics, FCT NOVA, Portugal
[4]Department of Economics and Statistics, University of Mauritius, Mauritius

Abstract In a system of regression models, finding a feasible shrinkage is demanding since the covariance structure is unknown and cannot be ignored. On the other hand, specifying sub-space restrictions for adequate shrinkage is vital. This study proposes feasible shrinkage estimation strategies where the sub-space restriction is obtained from LASSO. Therefore, some feasible LASSO-based Stein-type estimators are introduced, and their asymptotic performance is studied. Extensive Monte Carlo simulation and a real-data experiment support the superior performance of the proposed estimators compared to the feasible generalized least-squared estimator.

Keywords Feasible generalized least squares estimator, LASSO, Preliminary test estimation, Seemingly unrelated regression models, Shrinkage estimation, Stein-type estimation

AMS 2010 subject classifications 62J07, 62F10, 62H12

DOI: 10.19139/soic-2310-5070-1589

1. Introduction

A multiple linear regression model's goal is to describe the behavior of a given research variable (say, response) in terms of a group of explanatory variables. In addition, it is reasonable that more than one multiple regression equation must be utilized for real-life problems. Likewise, each equation may explain a different economic phenomenon in a set of individual linear multiple regression equations. For example, in a country with 20 states where the economic goal is to determine the country's consumption pattern, each state has to be attributed with a specific consumption equation, which ultimately makes 20 consumption functions. The question of interest is the formulation of different multiple regression equations and the associated mathematical challenges.

The possible approach is to consider a set of simultaneous equations models associated with each other. It entails another possibility where none of the system's dependent variables are explanatory in different equations. However, interactions between the separate equations may still exist if at least some random error components associated with the various equations are correlated. Because of the joint relation of the distribution of error terms and the non-diagonal covariance matrix, these equations can be linked statistically but not structurally. The "seemingly unrelated regression equations" (SUR)[43] model shows this type of behavior, while the individual equations may not appear to be linked to one another at first glance. For more information about the linear SUR model, the reader may refer to Srivastava and Jiles

*Correspondence to: Mina Norouzirad (Email: mina.norouzirad@gmail.com, m.norouzirad@fct.unl.pt). Center for Mathematics and Applications (NovaMath), FCT NOVA, Portugal.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

[38]. Thus, a system regression is used to solve linear models that appear unrelated but are related. It may seem pointless at first since they could be solved separately. However, there may be correlations between error terms that can be used to improve estimates.

It may not be necessary to include the same variables in all models; variables in each equation may be different. While the equations appear distinct, there may still be some relationship between them. Such simultaneous equations can determine the jointness of the distribution of the disturbances. Likewise, demand equations are frequently specified in-demand studies to explain household-level consumption for several commodities. The correlation between the equation disturbances could be due to various factors, including correlated shocks to household income [5].

Now, consider the M-system of regression models given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_M^\top)^\top$, $\boldsymbol{X} = \mathrm{Diag}(\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_\mathrm{M})$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_M)^\top$, and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_M)^\top$, in which $\boldsymbol{y}_k = (y_{k.1}, \ldots, y_{k.n_k})^\top$ is an n_k-dimensional vector of dependents variables. For the number of observations, n_k and the number of parameters, p_k, $\boldsymbol{X}_k = (X_{k.ij}), i = 1, \ldots, n_k; j = 1, \ldots, p_k$ is the matrix of explanatory variables. Finally, $\boldsymbol{\beta}_k = (\beta_{k.1}, \ldots, \beta_{k.p_k})^\top$ is a p_k-dimensional vector of unknown parameters, and $\boldsymbol{\varepsilon}_k = (\varepsilon_{k.1}, \ldots, \varepsilon_{k.n_k})^\top$ is an n_k-dimensional vector of errors in the kth equation, with $k = 1, \ldots, M$.

Suppose that $n^* = \sum_{k=1}^{M} n_k$ and $p = \sum_{k=1}^{M} p_i$. As a result, $\boldsymbol{y} \sim n^* \times 1$, $\boldsymbol{X} \sim n^* \times p$, $\boldsymbol{\beta} \sim p \times 1$, and $\boldsymbol{\varepsilon} \sim n^* \times 1$ are all easily found. Sometimes in practical situations, the number of observations in all equations is the same, i.e., $\forall k, n_k = n$, in this case $n^* = nM$. We avoid confusing the reader throughout the paper by using multiple notations with different subscripts. Thus, variables read as $\boldsymbol{y} \sim nM \times 1$, $\boldsymbol{X} \sim nM \times p$, and $\boldsymbol{\varepsilon} \sim nM \times 1$. Although it is preferable to use the notations $\boldsymbol{y}_n$, $\boldsymbol{X}_n$, and $\boldsymbol{\varepsilon}_n$ rather than $\boldsymbol{y}$, $\boldsymbol{X}$, and $\boldsymbol{\varepsilon}$, we disregard it due to a misunderstanding of notations.

For the stacked errors, considering the interactions between the M equations, we shall assume: (i) $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{X}_i^\top \boldsymbol{X}_j = \boldsymbol{C}_{ij}$, where $\boldsymbol{C}_{ij}$ is a non-singular matrix with fixed and finite components, (ii) The $nM \times nM$ covariance matrix is composed of n^2 blocks of M^2-blocks of the form $\mathrm{E}(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j) = \sigma_{ij}^2 \boldsymbol{I}_n$ where σ_{ij}^2 is the covariance between the disturbances of ith and jth equations for each observation in the sample. More succinctly, $\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$, and $\mathrm{E}[\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^\top] = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n = \boldsymbol{\Omega}$ where $\otimes$ is the Kronecker product. The matrix $\boldsymbol{\Omega}$ is of dimension $nM \times nM$, and $\boldsymbol{\Sigma} = (\sigma_{ij}^2)$ is an unstructured positive definite symmetric matrix of dimension $M \times M$. Thus, there are no linear relationships between the contemporaneous disturbances in the M equations of the model.

The structure of $\boldsymbol{\Omega}$ implies that n errors in each of the M equations have zero mean, equal variance, and are uncorrelated, i.e., $\mathrm{E}\left[\boldsymbol{\varepsilon}_{k.i} \boldsymbol{\varepsilon}_{k.j}\right] = 0$ for all $i, j = 1, \ldots, n$. The covariance between contemporaneous errors for a pair of equations are potentially nonzero but equal, while non-contemporaneous/intertemporal covariance are all zero, i.e.,

$$\mathrm{E}\left[\boldsymbol{\varepsilon}_{i.t} \boldsymbol{\varepsilon}_{j.s}\right] = 0, \quad \mathrm{E}\left[\boldsymbol{\varepsilon}_{i.t} \boldsymbol{\varepsilon}_{j.t}\right] = \sigma_{ij}^2 \qquad \forall t, s = 1, \ldots, n; \quad \forall i, j = 1, \ldots, M.$$

Although the terms "contemporaneous" and "intertemporal" covariance imply that data are available in time-series form, this is not required. Also, it is helpful for cross-sectional data. The homoskedastic disturbances in a single equation model are natural generalizations of contemporaneous covariance.

In situations where we are provided with some prior information in the form of a system of equations on $\boldsymbol{\beta}$, it is well-known that incorporating such information in estimation will improve the results. See Saleh [30] Saleh et al. [27, 29]. However, it is also challenging to integrate such information into the estimation. In this paper, the shrinkage estimation strategy attempts to incorporate general uncertain prior information (here, $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$) into the estimation process, where we use the least absolute shrinkage and selection operator (LASSO) of Tibshirani [40] for eliciting the prior information. We consider three critical members of this family: (1) the preliminary test estimator (PTE), (2) the Stein-type estimator (STE), and (3) the positive-part Stein-type estimator (PRSTE).

Many researchers have looked into these types of $\boldsymbol{\beta}$ estimations in different models. Saleh [30] is one of the references. The interested readers are referred to Kleyn et al. [14] for economy; to Saleh and Norouzirad [32] for non-orthogonal regression models, to Norouzirad et al. [21, 23], Norouzirad and Arashi [19, 20], Saleh et al. [28, 31] for robust models; to Arashi et al. [7], Norouzirad et al. [22] for partially linear models; to Roozbeh[24], Arashi and Roozbeh [6] for semi-parametric model, to NooriAsl et al. [18] for Poisson regression model; to Zandi et al. [42] for different types of mixed models; to Mandal et al.[16] Mahmoudi et al. [15] for gamma regression models; to Kashani [13] for fuzzy regression models; to Safariyan et al.[25, 26] for estimating of stress-strength reliability based on ranked set sampling; to Arashi and Tabatabaey [8], Arashi et al. [3, 4] for multivariate elliptic models; and etc.

The rest of this paper is organized as follows. The feasible generalized least squares estimator is reviewed in Section 2. In Section 3, the shrinkage estimators are introduced. Section 4 focuses on the properties of different estimators, which are investigated through the asymptotic. A numerical study including a Monte Carlo simulation and Fringe dataset for investigating the performance of estimators is given in Section 5. Some conclusive remarks are presented in section 6.

2. Feasible Generalized Least Squares

Because there are no simultaneous variables in the system, each equation has its explanatory variables to explain the response. The M equations appear to be unrelated. The model's equations are linked stochastically via the serial correlation of disturbances across the model's equations. As a result, the SUR model is the name given to this system.

Each of the M equations is assumed to satisfy the linear regression model's classical assumptions and can be estimated separately. Naturally, this method ignores the correlation between the errors of different equations, which can be utilized by joint estimation. Zellner [43] introduced a generalized least squares (GLS) estimator for estimating the coefficients of a set of SUR models as follows:

$$\hat{\boldsymbol{\beta}}^{\text{GLS}} = \left[\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X}\right]^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{y}, \tag{2}$$

where $\boldsymbol{\Omega}^{-1} = \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n$. $\text{E}[\hat{\boldsymbol{\beta}}^{\text{GLS}}] = \boldsymbol{\beta}$ in this case, and $\text{Var}(\hat{\boldsymbol{\beta}}^{\text{GLS}}) = \left(\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X}\right)^{-1}$.

Assuming $\boldsymbol{D} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top - \left(\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega}^\top$, by the positive definiteness of $\boldsymbol{\Omega}$, $\text{Var}(\hat{\boldsymbol{\beta}}^{\text{LS}}) - \text{Var}(\hat{\boldsymbol{\beta}}^{\text{GLS}}) = \boldsymbol{D}\boldsymbol{\Omega}\boldsymbol{D}^\top$ is at least a positive semi-definite matrix where $\hat{\boldsymbol{\beta}}^{\text{LS}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$. Thus, for estimating $\boldsymbol{\beta}$, the GLS estimator is more efficient than the ordinary least squares estimator. See Fiebig [9], Srivastava and Dwivedi [37], Srivastava and Jiles [38] for a succinct review of the literature in this field.

However, since $\boldsymbol{\Sigma}$ is unknown, the estimator in (2) is not practical. The feasible generalized least squares (FGLS) estimators are investigated by substituting a consistent estimator (say, $\boldsymbol{S}$) of $\boldsymbol{\Sigma}$ in Eq. (2). Therefore, the FGLS estimator of $\boldsymbol{\beta}$ is obtained from

$$\hat{\boldsymbol{\beta}}^{\text{FGLS}} = \left[\boldsymbol{X}^\top \hat{\boldsymbol{\Omega}}^{-1} \boldsymbol{X}\right]^{-1} \boldsymbol{X} \hat{\boldsymbol{\Omega}}^{-1} \boldsymbol{y}. \tag{3}$$

where $\hat{\boldsymbol{\Omega}} = \boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n$. Assume $\boldsymbol{S} = (s_{ij}^2)$ is a non-singular matrix and s_{ij}^2 is an estimator for σ_{ij}^2.

Due to Shalabh [35], there are two possible methods for estimating σ_{ij}^2:

(1) Consider $\boldsymbol{Z} = (\boldsymbol{X}_1 \vdots \boldsymbol{X}_2 \vdots \ldots \vdots \boldsymbol{X}_M)$ as an aggregated matrix with dimension $n \times p$, where $p = \sum_{i=1}^{M} p_i$. if we regress each of the M response on the columns of the Z, we get an $n \times 1$ residual vectors with the formula $\hat{\varepsilon}_k = \bar{\boldsymbol{P}}_Z y_k$, for $k = 1, \ldots, M$ where $\bar{\boldsymbol{P}}_Z = \boldsymbol{I}_n - \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top$. As a result, $s_{ij}^2 = \frac{1}{n} \hat{\varepsilon}_i \hat{\varepsilon}_j = \frac{1}{n} \boldsymbol{y}_i^\top \bar{\boldsymbol{P}}_Z \boldsymbol{y}_j$. It is simple to demonstrate that $\frac{n}{n-p} s_{ij}^2$ is an unbiased estimator for σ_{ij}^2.

(2) Regress each equation, i.e, regress $\boldsymbol{y}_k$ on $\boldsymbol{X}_k$ using the OLS method for $k = 1, \ldots, M$, and find the residual vector as $\tilde{\varepsilon}_k = \bar{\boldsymbol{H}}_X y_k$ where $\bar{\boldsymbol{H}}_X = \boldsymbol{I}_n - \boldsymbol{X}_k(\boldsymbol{X}_k^\top \boldsymbol{X}_k)^{-1}\boldsymbol{X}_k^\top$. A consistent estimator of σ_{ij}^2 is obtained by $s_{ij}^{*^2} = \frac{1}{n}\tilde{\varepsilon}_i\tilde{\varepsilon}_j = \frac{1}{n}\boldsymbol{y}_i^\top \bar{\boldsymbol{H}}_{X_i}\bar{\boldsymbol{H}}_{X_j}\boldsymbol{y}_j$. Thus, to find an unbiased estimate of σ_{ij}^2 based on $s_{ij}^{*^2}$, replace n with $\mathrm{tr}(\bar{\boldsymbol{H}}_{X_i}\bar{\boldsymbol{H}}_{X_j}) = n - p_i - p_j + \mathrm{tr}((\boldsymbol{X}_i^\top \boldsymbol{X}_i)^{-1}\boldsymbol{X}_i^\top \boldsymbol{X}_j(\boldsymbol{X}_j^\top \boldsymbol{X}_j)^{-1}\boldsymbol{X}_j^\top \boldsymbol{X}_i$.

Several approaches have been proposed to improve the FGLS estimator. Using extraneous or prior information can help estimators improve. In applied research, prior information about the regression coefficients may be available. According to the constant returns to scale in economics, the exponents in a Cobb-Douglas production function should sum to one for the function to be considered valid. In another example, consumers' lack of belief in the existence of money implies that the sum of money income and the price elasticized in a demand function should be zero. It is possible to obtain information about these constraints or prior knowledge from theoretical considerations, the experimenter's prior experience, empirical research, outside sources, etc.

Assume that the prior information that binds the regression coefficients is available from some extraneous source and can be expressed in the form of exact linear constraints as follows:

$$\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h},$$

when $\boldsymbol{H}$ is a known $q \times p$ matrix of rank q $(< p)$ and $\boldsymbol{h}$ is a q-vector of known constants.

Sengupta and Jammalamadaka [34] argue that the restrictions may be (a) a fact known from theoretical or experimental considerations, (b) a hypothesis that may have to be tested, or (c) an artificially imposed condition to reduce or eliminate redundancy in the description of the model.

Assume the true regression parameter vector $\boldsymbol{\beta}$ can be partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2 = \boldsymbol{0}$ are of order p_1 and p_2, respectively with $p_1 + p_2 = p$. We can select the important variables per equations in the SUR model using LASSO. Let $\boldsymbol{\beta}_i = (\boldsymbol{\beta}_{i1}^\top, \boldsymbol{\beta}_{i2}^\top)^\top$; we can write $\boldsymbol{\beta}_1 = (\boldsymbol{\beta}_{11}, \ldots, \boldsymbol{\beta}_{M1})^\top \sim p_1 \times 1$ and $\boldsymbol{\beta}_2 = (\boldsymbol{\beta}_{12}, \ldots, \boldsymbol{\beta}_{M2})^\top \sim p_2 \times 1$. Thus, the LASSO will define a sub-model that only has $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2 = \boldsymbol{0}$. Following these considerations, this partition is a special case of a restriction, $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$ in case of $\boldsymbol{H} = [\boldsymbol{0} \quad \boldsymbol{I}]$ and $\boldsymbol{h} = \boldsymbol{0}$ where $\boldsymbol{0}$ is $p_2 \times p_1$ matrix of zeros and $\boldsymbol{I}$ is the identity matrix of order $p_2 \times p_2$. Here, the dimension of $\boldsymbol{H}$, q is equal to p_2.

Using both sample data and prior information simultaneously with the restricted least squares estimation method is possible. This method selects $\boldsymbol{\beta}$ such that the error sum of squares is minimized subject to $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$. This hypothesis is compared with the one in Arashi and Roozbeh [5].

The restricted FGLS (RFGLS) estimator can be found by

$$\hat{\boldsymbol{\beta}}^{\mathrm{RFGLS}} = \hat{\boldsymbol{\beta}}^{\mathrm{FGLS}} - \left(\boldsymbol{X}^\top \hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{H}^\top \left(\boldsymbol{H}(\boldsymbol{X}^\top \hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{H}^\top\right)^{-1}\left(\boldsymbol{H}\hat{\boldsymbol{\beta}}^{\mathrm{FGLS}} - \boldsymbol{h}\right). \tag{4}$$

Let us define $\boldsymbol{G}_n^1 = (\boldsymbol{X}^\top \hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X})^{-1}$ and $\boldsymbol{G}_n^2 = (\boldsymbol{H}\boldsymbol{G}_n^1\boldsymbol{H}^\top)^{-1}$. It is worth noting that $\boldsymbol{G}_n^1$ is a positive definite matrix and the FGLS estimator's variance-covariance matrix. Now, we can rewrite the equations (3) and (4) as

$$\hat{\boldsymbol{\beta}}^{\mathrm{FGLS}} = \boldsymbol{G}_n^1\boldsymbol{X}^\top \hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{y}, \tag{5}$$

$$\hat{\boldsymbol{\beta}}^{\mathrm{RFGLS}} = \hat{\boldsymbol{\beta}}^{\mathrm{FGLS}} - \boldsymbol{G}_n^1\boldsymbol{H}^\top \boldsymbol{G}_n^2\left(\boldsymbol{H}\hat{\boldsymbol{\beta}}^{\mathrm{FGLS}} - \boldsymbol{h}\right). \tag{6}$$

Because $\boldsymbol{H}\boldsymbol{\beta} \neq \boldsymbol{h}$ in this case, the FGLS estimator does not follow the restriction $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$ and thus does not use the prior information; so, it is known as an unrestricted FGLS estimator (UFGLSE). As a result, the question is how to use the information from the samples with prior knowledge to produce an improved estimator of $\boldsymbol{\beta}$. We propose utilizing the shrinkage approach for SUR models to answer this question.

Shrinkage estimation has established itself as an essential technique for data modeling, attempting to attract the interest of academics and practitioners in a wide range of fields. This estimation technique enables the combination of data from different sources. Zellner and Vandaele [44] proposed shrinkage estimations in SUR, which extends the results of James and Stein [11] and Sclove [33] to multivariate regression equations and gives a technique for creating an estimate whose risk is less than the risk of the GLS estimator. However, the resulting estimator is not practical because it depends on unknown matrices. Srivastava [36] investigates the estimator's properties when these unknown matrices are replaced with consistent estimators.

Srivastava and Wan [39] proposed the following formula for a general Stein-rule estimator of $\boldsymbol{\beta}$:

$$\tilde{\boldsymbol{\beta}}^{\mathrm{SW}} = \left[1 - \left(\frac{h}{2n-p+2}\right)\frac{(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathrm{FGLS}})^{\top}\hat{\boldsymbol{\Omega}}^{-1}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathrm{FGLS}})}{\hat{\boldsymbol{\beta}}^{\mathrm{FGLS}\top}\boldsymbol{X}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathrm{FGLS}\top}}\right]\hat{\boldsymbol{\beta}}^{\mathrm{FGLS}}, \tag{7}$$

where h is a positive non-stochastic shrinkage factor characterizing the estimator.

Mehrabani and Ullah [17] consider a Stein-type shrinkage as an averaging estimator, which is a weighted average of the unrestricted GLS and the restricted GLS under the constraint $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \ldots = \boldsymbol{\beta}_k = \boldsymbol{\beta}$. The weight has an inverse relationship with a weighted quadratic loss function, which quantifies the weighted difference between the unrestricted and restricted GLS estimators. They defined the average estimator as $\hat{\boldsymbol{\beta}}_A = (1-\frac{\tau}{D})\hat{\boldsymbol{\beta}} + \frac{\tau}{D}\tilde{\boldsymbol{\beta}}$ where $D = (\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})^{\top}\boldsymbol{W}(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})$ is a quadratic loss function with W, an arbitrary symmetric positive definite weight matrix, and τ, a positive characterizing parameter.

Yuzbasi and Ahmed [41] assumed that each equation's vector regression coefficient in a SUR model has two parts: one for the main effects and another for the nuisance effects, which may be close to zero. Thus, each equation was subjected to two competing models: one that included all coefficient regressions (full model) and another that included only the coefficients of the main effects based on the auxiliary information (sub-model). Considering explanatory variables are affected by multicollinearity, they proposed restricted (or sub-model) ridge, preliminary, and shrinkage SUR regression estimations to improve the full model ridge SUR regression [1].

3. Shrinking towards the prior information

Under the assumption that the model has been correctly specified, it is common practice to conduct a statistical model on a sample of data. Some estimators (e.g. OLS, maximum likelihood, Bayesian, ...) have been used for the purposes of estimation and drawing statistical inferences from the sample. Sometimes non-sample information provides prior information or constraints. For efficient estimation, it is important to incorporate and use such prior knowledge, provided it is correct. As a result, the suitability of the estimator depends on the accuracy of the underlying assumptions. To remove the uncertainty in the null hypothesis, $\mathcal{H}_o : \boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h}$, the Wald test is used. It is defined by

$$\mathcal{W} = \frac{1}{n}\left(\boldsymbol{H}\hat{\boldsymbol{\beta}}^{\mathrm{FGLS}} - \boldsymbol{h}\right)^{\top}\boldsymbol{G}_n^2\left(\boldsymbol{H}\hat{\boldsymbol{\beta}}^{\mathrm{FGLS}} - \boldsymbol{h}\right). \tag{8}$$

Under the null hypothesis, $\mathcal{W}$ has a χ^2 distribution with q degree of freedom (d.f.) [10, p. 347].

3.1. Shrinkage estimators

From now on, let us use the notations $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ instead of $\hat{\boldsymbol{\beta}}^{\mathrm{FGLS}}$ and $\hat{\boldsymbol{\beta}}^{\mathrm{RFGLS}}$ to avoid confusion in the notations of the following introduced estimators.

The family of shrinkage estimators of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}^{\mathrm{Shrinkage}} = g(\mathcal{W})\hat{\boldsymbol{\beta}} + (1-g(\mathcal{W}))\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} - g(\mathcal{W})(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}). \tag{9}$$

where $\mathcal{W}$ is the test-statistic for testing the null hypothesis, and $g(\mathcal{W})$ is a non-decreasing function of $\mathcal{W}$.

In this study, $g(\mathcal{W})$ is limited to the following functions:

(i) $g(\mathcal{W}) = 0$, then $\hat{\boldsymbol{\beta}}^{\mathrm{Shrinkage}} = \tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{\mathrm{FGLS}}$, the unrestricted FGLS estimator (or briefly, FGLS estimator).

(ii) $g(\mathcal{W}) = 1$, then $\hat{\boldsymbol{\beta}}^{\text{Shrinkage}} = \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{\text{RFGLS}}$, the restricted FGLS estimator (or briefly, restricted estimator).

(iii) $g(\mathcal{W}) = I(\mathcal{W} \leq \chi^2_{q,\alpha})$, then $\hat{\boldsymbol{\beta}}^{\text{Shrinkage}} = \hat{\boldsymbol{\beta}}^{\text{PT}}(\alpha)$, the preliminary test FGLS (PTFGLS) estimator (or briefly, preliminary test estimator).

(iv) $g(\mathcal{W}) = (q-2)\mathcal{W}^{-1}$, then $\hat{\boldsymbol{\beta}}^{\text{Shrinkage}} = \hat{\boldsymbol{\beta}}^{\text{S}}$, the Stein-type shrinkage FGLS (SFGLS) estimator (or briefly, Stein-type estimator).

(iv) $g(\mathcal{W}) = (q-2)\mathcal{W}^{-1} + \left(1 - (q-2)\mathcal{W}^{-1}\right) I(\mathcal{W} \leq (q-2))$, then $\hat{\boldsymbol{\beta}}^{\text{Shrinkage}} = \hat{\boldsymbol{\beta}}^{\text{S+}}$, the positive part Stein-type shrinkage FGLS (PRSFGLS) estimator (or briefly, positive rule estimator).

To avoid confusion, Table 1 summarizes the shrinkage estimators.

Table 1. The shrinkage family member's

$g(\mathcal{W})$	Notation	Name	Abbr.	A.K.A.
0	$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{\text{FGLS}}$	Unrestricted FGLS	UFGLS	FGLS
1	$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{\text{RFGLS}}$	Restricted FGLS	RFGLE	Restricted
$I(\mathcal{W} \leq \chi^2_{q,\alpha})$	$\hat{\boldsymbol{\beta}}^{\text{PT}}$	Preliminary test FGLS	PTFGLS	Preliminary test
$(q-2)\mathcal{W}^{-1}$	$\hat{\boldsymbol{\beta}}^{\text{S}}$	Stein-type shrinkage FGLS	SFGLS	Stein-type
$(q-2)\mathcal{W}^{-1} + \left(1-(q-2)\mathcal{W}^{-1}\right) I(\mathcal{W} \leq (q-2))$	$\hat{\boldsymbol{\beta}}^{\text{S+}}$	Positive part Stein-type shrinkage FGLS	PRSFGLS	Positive rule

3.1.1. Preliminary test estimation When the true sampling model's content is unknown, the current statistical model can be determined by a preliminary test of hypothesis using the available sample data. These procedures are carried out in two stages and are based on a hypothesis test that establishes a rule for selecting between two estimators: (1) based on sample data and (2) consistent with the hypothesis. This necessitates the conduct of a compatibility test between the FGLS estimator purely based on sample information and the restricted estimator based on the linear hypothesis. Depending on the outcome, one can choose an estimator. Thus, the FGLS or restricted estimator can be selected. This procedure is called the preliminary test FGLS estimate of $\boldsymbol{\beta}$ and is defined as

$$\hat{\boldsymbol{\beta}}^{\text{PT}} = \tilde{\boldsymbol{\beta}} - \left(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}\right) I\left(\mathcal{W} \leq \chi^2_{q,\alpha}\right), \tag{10}$$

where $\chi^2_{q,\alpha}$ denotes $(1-\alpha)$th quantile of a χ^2_q variable such that $100(1-\alpha)\%$ area under the curve of the χ^2 distribution is to the left of the χ^2 distribution with q d.f., $\chi^2_{q,\alpha}$, for large n.

If $\mathcal{H}_o$ is accepted at level α of significance, the prior information is correct, so the restricted FGLS estimator is better than FGLS estimation; the preliminary test estimation improves the estimation procedure. On the other hand, if $\mathcal{H}_o$ is rejected at the α significance level, we conclude that the FGLS estimator is superior to the restricted estimator.

3.1.2. Stein-type shrinkage estimator It is worth noting that the probability of type I error (rejecting $\mathcal{H}_o$ when it is true) is 0 for $\alpha = 0$ and 1 for $\alpha = 1$. As a result, the entire area under the sampling distribution represents the area of acceptance or rejection of the null hypothesis. Therefore, preliminary test estimators' sampling performance is heavily dependent on their α choice.

Further, for $q \geq 3$ a Stein-type shrinkage FGLS estimator is defined as follows:

$$\hat{\boldsymbol{\beta}}^{\text{S}} = \tilde{\boldsymbol{\beta}} - \frac{q-2}{\mathcal{W}}\left(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}\right); \qquad q \geq 3. \tag{11}$$

The idea behind this estimator is that when the test statistic, $\mathcal{W}$ defined in Eq. (8) is small, the Stein-type estimator gives a higher weight to the restricted estimator, as it is the most efficient estimator. Nonetheless, when the value of $\mathcal{W}$ is large, the Stein-type estimator shrinks the FGLS estimator in the direction of the restricted estimator and always has the remarkable property of yielding a lower risk than the FGLS estimator provided with $p \geq 3$.

3.1.3. Positive-rule Stein-type shrinkage estimator The Stein-type estimator is not a convex combination of the restricted and unrestricted FGLS estimators. As a result, this estimator has the potential to change the sign of the FGLS estimator. To prevent this strange behavior, we truncate the SFGLS estimator, resulting in a convex combination of the FGLS and restricted estimators. We call this a positive-rule Stein-type FGLS estimator. This estimator is denoted by $\hat{\boldsymbol{\beta}}^{\mathrm{S}+}$ and defined as

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{\mathrm{S}+} &= \hat{\boldsymbol{\beta}} + \left(1 - \frac{q-2}{\mathcal{W}}\right) I(\mathcal{W} > q-2)(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) \\ &= \hat{\boldsymbol{\beta}}^{\mathrm{S}} - \left(1 - \frac{q-2}{\mathcal{W}}\right) I\left(\mathcal{W} \leq q-2\right)\left(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}\right); \quad q \geq 3. \end{aligned} \tag{12}$$

4. Asymptotic Bias and Risk Performance

This section derives some fundamental characteristics of the estimators that we discussed previously.

In order to provide a reasonable analysis, a sequence of local alternatives $\mathcal{K}_n$ is taken into consideration and is represented by the expression

$$\mathcal{K}_n : \boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{h} + \frac{\boldsymbol{\xi}}{\sqrt{n}}, \quad \boldsymbol{\xi} = (\xi_1, \dots, \xi_q)^\top.$$

For our purpose, we need to set the following regularity conditions.

Here, for each of M equations as the classical regression model, the following assumptions are satisfied: (i) $\boldsymbol{X}_k$ is fixed. (ii) $p_k = \mathrm{rank}(\boldsymbol{X}_k)$, (iii) $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{X}_k^\top \boldsymbol{X}_k = \boldsymbol{C}_{kk}$ where $\boldsymbol{C}_{kk}$ is a non-singular matrix with fixed and finite elements, (iv) $\mathrm{E}\left[\boldsymbol{\varepsilon}_k\right] = 0$, (v) $\mathrm{E}\left[\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k^\top\right] = \sigma_{kk}^2 \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is an identity matrix of $n \times n$ dimension, and σ_{kk}^2 is the variance of disturbances in the kth equation.

(i) $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{G}_n^1 = \boldsymbol{G}_1, \quad \boldsymbol{G}_1 = (\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}$,
(ii) $\lim_{n\to\infty} n\boldsymbol{G}_n^2 = \boldsymbol{G}_2 \quad \boldsymbol{G}_2 = (\boldsymbol{H}^\top\boldsymbol{G}^1\boldsymbol{H})^{-1}$.

Amemiya [2, p. 198] demonstrated that the FGLS estimator is unbiased for a fixed n by assuming that the error terms have a symmetric distribution. In large samples, it is consistent and asymptotically normal with the limiting distribution,

$$\sqrt{n}\left(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{G}_1), \tag{13}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. Note that in small samples, it does not follow normal distribution.

Allow $\boldsymbol{Q}_1$ to equal $\sqrt{n}\left(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)$. According to Eq. (13), it has asymptotic distribution $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{G}_1)$. Using this fact, under $\{\mathcal{K}_n\}$, we have

$$\begin{aligned} \boldsymbol{Q}_2 &= \boldsymbol{H}\boldsymbol{Q}_1 + \boldsymbol{\xi} = \sqrt{n}\left(\boldsymbol{H}\tilde{\boldsymbol{\beta}} - \boldsymbol{h}\right) \xrightarrow{\mathcal{D}} \mathcal{N}_q\left(\boldsymbol{\xi}, \boldsymbol{G}_2^{-1}\right), \\ \boldsymbol{Q}_3 &= \boldsymbol{Q}_1 - \boldsymbol{G}_1\boldsymbol{H}^\top\boldsymbol{G}_2\boldsymbol{Q}_2 = \sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{\mathcal{D}} \mathcal{N}_p\left(-\boldsymbol{\delta}, \boldsymbol{G}_1 - \boldsymbol{G}^\star\right), \\ \boldsymbol{Q}_4 &= \boldsymbol{G}_1\boldsymbol{H}^\top\boldsymbol{G}_2\boldsymbol{Q}_2 = \sqrt{n}\left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}_p\left(\boldsymbol{\delta}, \boldsymbol{G}^\star\right), \end{aligned}$$

where $\boldsymbol{\delta} = \boldsymbol{G}_1\boldsymbol{H}^\top\boldsymbol{G}_2\boldsymbol{\xi}$, and $\boldsymbol{G}^\star = \boldsymbol{G}_1\boldsymbol{H}^\top\boldsymbol{G}_2\boldsymbol{H}\boldsymbol{G}_1$, and q is the number of rows in matrix $\boldsymbol{H}$. It is easy to conclude the following lemma.

Lemma 4.1
Under $\mathcal{K}_n$, we have

$$\begin{aligned}
\begin{bmatrix}\boldsymbol{Q}_1\\ \boldsymbol{Q}_2\end{bmatrix} &\xrightarrow{\mathcal{D}} \mathcal{N}_{p+q}\left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{\xi}\end{bmatrix}, \begin{bmatrix}\boldsymbol{G}_1 & \boldsymbol{G}_1\boldsymbol{H}^\top\\ \boldsymbol{H}\boldsymbol{G}_1 & \boldsymbol{G}_2^{-1}\end{bmatrix}\right),\\
\begin{bmatrix}\boldsymbol{Q}_1\\ \boldsymbol{Q}_4\end{bmatrix} &\xrightarrow{\mathcal{D}} \mathcal{N}_{2p}\left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{\delta}\end{bmatrix}, \begin{bmatrix}\boldsymbol{G}_1 & \boldsymbol{G}^\star\boldsymbol{H}^\top\\ \boldsymbol{G}^\star & \boldsymbol{G}^\star\end{bmatrix}\right),\\
\begin{bmatrix}\boldsymbol{Q}_3\\ \boldsymbol{Q}_4\end{bmatrix} &\xrightarrow{\mathcal{D}} \mathcal{N}_{2p}\left(\begin{bmatrix}-\boldsymbol{\delta}\\ \boldsymbol{\delta}\end{bmatrix}, \begin{bmatrix}\boldsymbol{G}_1-\boldsymbol{G}^\star & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{G}^\star\end{bmatrix}\right).
\end{aligned}$$

Since $\mathcal{W} = \boldsymbol{Q}_2^\top(n^{-1}\boldsymbol{G}_n^2)\boldsymbol{Q}_2$, it will be equal to $\mathcal{W} = \boldsymbol{Q}_2^\top\boldsymbol{G}_2\boldsymbol{Q}_2$ as $n\to\infty$ by the regularity assumption (ii). Therefore, the asymptotic distribution of $\mathcal{W}$ under $\mathcal{K}_n$ is $\chi^2_q(\Delta)$, the non-central χ^2 distribution with q d.f. and the non-centrality parameter $\Delta = \boldsymbol{\delta}^\top\boldsymbol{G}_2\boldsymbol{\delta}$.

We need the following lemma to find the asymptotic distributional bias and risk of the proposed estimators.

Lemma 4.2
Let $\boldsymbol{Z} = (Z_1,\ldots,Z_k)^\top$ be a k-dimensional vector with the distribution $\mathcal{N}_k(\boldsymbol{\mu},\boldsymbol{I}_k)$. For a measurable function ϕ, we have

$$\begin{aligned}
\mathrm{E}\left[\boldsymbol{Z}\phi\left(\boldsymbol{Z}^\top\boldsymbol{Z}\right)\right] &= \boldsymbol{\mu}\,\mathrm{E}\left[\phi\left(\chi^2_{k+2}(\Delta)\right)\right],\\
\mathrm{E}\left[\boldsymbol{Z}\boldsymbol{Z}^\top\phi(\boldsymbol{Z}^\top\boldsymbol{Z})\right] &= \boldsymbol{I}_p\,\mathrm{E}\left[\phi\left(\chi^2_{k+2}(\Delta)\right)\right] + \boldsymbol{\mu}\boldsymbol{\mu}^\top\,\mathrm{E}\left[\phi\left(\chi^2_{k+4}(\Delta)\right)\right].
\end{aligned}$$

Proof
See [12]. □

4.1. Asymptotic distributional Bias (ADB) Performance

The asymptotic distributional bias (ADB) of an arbitrary estimator $\hat{\boldsymbol{\beta}}^\star$ is defined as

$$\mathrm{ADB}(\hat{\boldsymbol{\beta}}^\star) = \mathrm{E}\left[\lim_{n\to\infty}\sqrt{n}(\hat{\boldsymbol{\beta}}^\star - \boldsymbol{\beta})\right]. \tag{14}$$

Theorem 4.1
Suppose that the assumptions (i) and (ii) are held. Under $\mathcal{K}_n$ and Eq. (13), the ADBs of FGLS, RFGLS, PTFGLS, SFGLS, and PRSFGLS estimators are given by

$$\begin{aligned}
&\mathrm{ADB}\left(\tilde{\boldsymbol{\beta}}\right) = \boldsymbol{0},\\
&\mathrm{ADB}\left(\hat{\boldsymbol{\beta}}\right) = -\boldsymbol{\delta}, \qquad \boldsymbol{\delta} = \boldsymbol{G}_1\boldsymbol{H}^\top\boldsymbol{G}_2(\boldsymbol{H}\boldsymbol{\beta}-\boldsymbol{h}),\\
&\mathrm{ADB}\left(\hat{\boldsymbol{\beta}}^{\mathrm{PT}}\right) = -\boldsymbol{\delta}H_{q+2}\left(\chi^2_{q,\alpha};\Delta\right),\\
&\mathrm{ADB}\left(\hat{\boldsymbol{\beta}}^{\mathrm{S}}\right) = -(q-2)\boldsymbol{\delta}\,\mathrm{E}\left[\chi^{-2}_{q+2}(\Delta)\right],\\
&\mathrm{ADB}\left(\hat{\boldsymbol{\beta}}^{\mathrm{S}+}\right) = \mathrm{ADB}\left(\hat{\boldsymbol{\beta}}^{\mathrm{S}}\right) - \boldsymbol{\delta}\,\mathrm{E}\left[\left(1-(q-2)\chi^{-2}_{q+2}(\Delta)\right)I\left(\chi^2_{q+2}(\Delta)\le(q-2)\right)\right],
\end{aligned}$$

where $\chi^2_\nu(\Delta)$ is the non-central χ^2 distribution with ν degrees of freedom and non-centrality parameter $\Delta = \boldsymbol{\delta}^\top\boldsymbol{G}_1^{-1}\boldsymbol{\delta} = n^{-1}\boldsymbol{\xi}^\top\boldsymbol{G}_2^\top\boldsymbol{H}\boldsymbol{G}_1\boldsymbol{H}^\top\boldsymbol{G}_2\boldsymbol{\xi}$ and $H_\nu(\cdot;\Delta)$ is its cumulative distribution function. Furthermore,

$$\begin{aligned}
\mathrm{E}\left[\chi_\nu^{-2}(\Delta)\right] &= \exp\left\{-\frac{\Delta}{2}\right\}\sum_{r\ge0}\frac{1}{r!}\left(\frac{\Delta}{r}\right)^r\frac{1}{\nu+2r-2} = \mathrm{E}_R\left(\nu+2r-2\right)^{-1},\\
\mathrm{E}\left[\chi_\nu^{-4}(\Delta)\right] &= \mathrm{E}_R\left[(\nu+2r-2)\left(\nu+2r-4\right)\right]^{-1},\\
\mathrm{E}\left[\chi_\nu^{-2}I\left(\chi^2_\nu(\Delta)\le c\right)\right] &= \mathrm{E}_R\left[(\nu+2r-2)^{-1}H_{\nu+2r-2}(c;0)\right],\\
\mathrm{E}\left[\chi_\nu^{-4}I\left(\chi^2_\nu(\Delta)\le c\right)\right] &= \mathrm{E}_R\left[(\nu+2r-2)^{-1}(\nu+2r-4)^{-1}H_{\nu+2r-4}(c;0)\right],
\end{aligned}$$

where E_R is the expectation of the variable (R) given by the Poisson distribution with parameter $\Delta/2$.

Proof
By the Eq. (14), definition of the ADB, under $\mathcal{K}_n$, we have

$$\begin{aligned} \mathrm{ADB}(\hat{\boldsymbol{\beta}}^{\text{Shrinkage}}) &= \lim_{n\to\infty} \mathrm{E}\left[\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{\text{Shrinkage}} - \boldsymbol{\beta}\right)\right] \\ &= \lim_{n\to\infty} \mathrm{E}\left[\sqrt{n}\left(\tilde{\boldsymbol{\beta}} - g(\mathcal{W})\left(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}\right) - \boldsymbol{\beta}\right)\right] \\ &= \lim_{n\to\infty} \mathrm{E}\left[\sqrt{n}\left(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right] - \lim_{n\to\infty} \mathrm{E}\left[\sqrt{n} g(\mathcal{W})\left(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}\right)\right] \\ &= \mathrm{ADB}(\tilde{\boldsymbol{\beta}}) - \lim_{n\to\infty} \mathrm{E}\left[\boldsymbol{Q}_4 g(\mathcal{W})\right] \\ &= \mathrm{ADB}(\tilde{\boldsymbol{\beta}}) - \lim_{n\to\infty} \mathrm{E}\left[\boldsymbol{G}_1 \boldsymbol{H}^\top \boldsymbol{G}_2 \boldsymbol{Q}_2\, g(\boldsymbol{Q}_2^\top \boldsymbol{G}_2 \boldsymbol{Q}_2)\right]. \end{aligned} \tag{15}$$

Using the Eq. (13), $\mathrm{ADB}(\tilde{\boldsymbol{\beta}}) = \boldsymbol{0}$ and considering $\boldsymbol{Z} = \boldsymbol{G}_2^{\frac{1}{2}} \boldsymbol{Q}_2$, we use Lemma 4.2 to obtain

$$\mathrm{ADB}(\hat{\boldsymbol{\beta}}^*) = -\boldsymbol{\delta}\, \mathrm{E}\left[g(\chi^2_{q+2}(\Delta^2))\right].$$

To complete the proof, just use the definition of shrinkage estimators. □

Additionally, the asymptotic distributional quadratic bias (ADQB) of an estimator $\hat{\boldsymbol{\beta}}^\star$ is given by

$$\mathrm{ADQB}\left(\hat{\boldsymbol{\beta}}^\star\right) = \left(\mathrm{ADB}\left(\hat{\boldsymbol{\beta}}^\star\right)\right)^\top \boldsymbol{G}_1 \left(\mathrm{ADB}\left(\hat{\boldsymbol{\beta}}^\star\right)\right).$$

Therefore, the ADQBs of the estimators are

$$\begin{aligned} \mathrm{ADQB}\left(\tilde{\boldsymbol{\beta}}\right) &= 0, \\ \mathrm{ADQB}\left(\hat{\boldsymbol{\beta}}\right) &= \Delta, \\ \mathrm{ADQB}\left(\hat{\boldsymbol{\beta}}^{\mathrm{PT}}\right) &= \Delta\left\{H_{q+2}\left(\chi^2_{q,\alpha};\Delta\right)\right\}^2, \\ \mathrm{ADQB}\left(\hat{\boldsymbol{\beta}}^{\mathrm{S}}\right) &= (q-2)^2 \Delta\left\{\mathrm{E}\left[\chi^{-2}_{q+2}(\Delta)\right]\right\}^2, \\ \mathrm{ADQB}\left(\hat{\boldsymbol{\beta}}^{\mathrm{S+}}\right) &= \Delta\left\{(q-2)\,\mathrm{E}\left[\chi^{-2}_{q+2}(\Delta)\right]\right. \\ &\quad \left. - \mathrm{E}\left[\left(1-(q-2)\chi^{-2}_{q+2}(\Delta)\right) I\left(\chi^2_{q+2}(\Delta) \le (q-2)\right)\right]\right\}^2. \end{aligned}$$

When $\Delta = 0$, the ADQB of all estimators is the same as that of the FGLS estimator. If $\Delta > 0$, the restricted estimator has no control over its ADQB. As $\Delta \to \infty$, $H_{q+2}\left(\chi^2_q(\alpha);\Delta\right) \to 0$. Using this fact, the ADQB of the preliminary test estimator vanishes when Δ approaches ∞, and it also offers good control of the ADQB function. The ADQB of the Stein-type estimator increases to a point and then decreases toward zero because $E\left[\chi^{-2}_{q+2}(\Delta)\right]$ is a non-decreasing log convex function of Δ. Finally, by comparing the ADQB function of the Stein-type and positive-rule estimators, it is obvious that the ADQB curve of the positive-rule estimator is the same or stays a little below the curve of the Stein-type estimator.

4.2. Asymptotic Distributional Risk (ADR) Performance

An estimator's asymptotic distributional risk (ADR) of an estimator $\hat{\boldsymbol{\beta}}^\star$ is given by

$$\mathrm{ADR}(\hat{\boldsymbol{\beta}}^\star) = \lim_{n\to\infty} \mathrm{E}\left[n(\hat{\boldsymbol{\beta}}^\star - \boldsymbol{\beta})^\top \boldsymbol{W} (\hat{\boldsymbol{\beta}}^\star - \boldsymbol{\beta})\right], \tag{16}$$

where $\boldsymbol{W}$ is a positive definite matrix.

Theorem 4.2
Under $\mathcal{K}_n$ and Eq. (13), the ADRs of FGLS, restricted, preliminary test, Stein-type, and positive-rule

FGLS estimators, respectively are given by

$$\begin{aligned}
\mathrm{ADR}\left(\tilde{\boldsymbol{\beta}}\right) &= \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{G}_1\right), \\
\mathrm{ADR}\left(\hat{\boldsymbol{\beta}}\right) &= \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{G}_1\right) - \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{G}^{\star}\right) + \boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}, \\
\mathrm{ADR}\left(\hat{\boldsymbol{\beta}}^{\mathrm{PT}}\right) &= \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{G}_1\right) - \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{G}^{\star}\right) H_{q+2}\left(\chi^2_{q,\alpha};\Delta\right) + \boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\, g_1(\alpha;\Delta), \\
\mathrm{ADR}\left(\hat{\boldsymbol{\beta}}^{\mathrm{S}}\right) &= \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{G}_1\right) - (q-2)\,\mathrm{tr}\left(\boldsymbol{W}\boldsymbol{G}^{\star}\right)\, g_2(\Delta) + (q^2-4)\boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\,\mathrm{E}\left[\chi^{-4}_{p+4}(\Delta)\right], \\
\mathrm{ADR}\left(\hat{\boldsymbol{\beta}}^{\mathrm{S+}}\right) &= \mathrm{ADR}\left(\hat{\boldsymbol{\beta}}^{\mathrm{S}}\right) - \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{G}^{\star}\right)\, g_3(\Delta) + \boldsymbol{\delta}^{\top}\boldsymbol{W}\boldsymbol{\delta}\, g_4(\Delta),
\end{aligned}$$

where $\boldsymbol{G}^{\star} = \boldsymbol{G}_1\boldsymbol{H}^{\top}\boldsymbol{G}_2\boldsymbol{H}\boldsymbol{G}_1$ and

$$\begin{aligned}
g_1(\alpha;\Delta) &= 2H_{q+2}(\chi^2_{q,\alpha};\Delta) - H_{q+4}(\chi^2_{q,\alpha};\Delta), \\
g_2(\Delta) &= 2\,\mathrm{E}\left[\chi^{-2}_{q+2}(\Delta)\right] - (q-2)\,\mathrm{E}\left[\chi^{-4}_{q+2}(\Delta)\right], \\
g_3(\Delta) &= 2\,\mathrm{E}\left[(1-(q-2)\chi^{-2}_{q+2}(\Delta))I(\chi^2_{q+2}(\Delta)\le(q-2))\right] \\
&\quad -2\,\mathrm{E}\left[(1-(q-2)\chi^{-2}_{q+4}(\Delta))I(\chi^2_{q+4}(\Delta)\le(q-2))\right] \\
&\quad -\mathrm{E}\left[(1-(q-2)\chi^{-2}_{q+4}(\Delta))^2 I(\chi^2_{q+4}(\Delta^2)\le(q-2))\right], \\
g_4(\Delta) &= H_{q+2}\left(q-2;\Delta\right) - (q-2)^2\,\mathrm{E}\left[\chi^{-4}_{q+2}(\Delta) I\left(\chi^2_{q+2}(\Delta)\le(q-2)\right)\right].
\end{aligned}$$

Proof

By the definition of the ADR, Eq. (16), we have

$$\mathrm{ADR}(\hat{\boldsymbol{\beta}}^{\mathrm{shrinkage}}) = \lim_{n\to\infty}\mathrm{E}\left[n\left(\hat{\boldsymbol{\beta}}^{\mathrm{shrinkage}}-\boldsymbol{\beta}\right)^{\top}\boldsymbol{W}\left(\hat{\boldsymbol{\beta}}^{\mathrm{shrinkage}}-\boldsymbol{\beta}\right)\right] = \mathrm{tr}\left(\boldsymbol{W}\boldsymbol{\Gamma}\right), \tag{17}$$

where $\boldsymbol{\Gamma} = \lim_{n\to\infty}\mathrm{E}\left[n\left(\hat{\boldsymbol{\beta}}^{\mathrm{shrinkage}}-\boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}}^{\mathrm{shrinkage}}-\boldsymbol{\beta}\right)^{\top}\right]$. As a result,

$$\begin{aligned}
\boldsymbol{\Gamma} &= \lim_{n\to\infty}\mathrm{E}\left[n\left(\tilde{\boldsymbol{\beta}} - g(\mathcal{W})(\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}\right)\left(\tilde{\boldsymbol{\beta}} - g(\mathcal{W})(\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}\right)^{\top}\right] \\
&= \lim_{n\to\infty}\mathrm{E}\left[\left(\boldsymbol{Q}_1 - g(\mathcal{W})\boldsymbol{Q}_4\right)\left(\boldsymbol{Q}_1 - g(\mathcal{W})\boldsymbol{Q}_4\right)^{\top}\right] \\
&= \lim_{n\to\infty}\mathrm{E}\left[\boldsymbol{Q}_1\boldsymbol{Q}_1^{\top}\right] - 2\lim_{n\to\infty}\mathrm{E}\left[\boldsymbol{Q}_1^{\top}\boldsymbol{Q}_4 g(\mathcal{W})\right] + \lim_{n\to\infty}\mathrm{E}\left[\boldsymbol{Q}_4\boldsymbol{Q}_4^{\top} g^2(\mathcal{W})\right].
\end{aligned} \tag{18}$$

The first term is the $\boldsymbol{Q}_1$ covariance matrix of which equals $\boldsymbol{G}_1^{-1}$. To calculate the second term, we need to know the asymptotically multivariate normal conditional distribution of $\boldsymbol{Q}_1|\boldsymbol{Q}_2$, which has a mean vector $\boldsymbol{G}_1\boldsymbol{H}^{\top}\boldsymbol{G}_2\left(\boldsymbol{Q}_2-\boldsymbol{\delta}\right)$ and covariance matrix $\boldsymbol{G}_1 - \boldsymbol{G}^{\star}$. Using this fact, we have,

$$\begin{aligned}
\lim_{n\to\infty}\mathrm{E}\left[\boldsymbol{Q}_1^{\top}\boldsymbol{Q}_4 g(\mathcal{W})\right] &= \boldsymbol{G}_1\boldsymbol{H}^{\top}\boldsymbol{G}_2\lim_{n\to\infty}\mathrm{E}\left[\boldsymbol{Q}_1^{\top}\boldsymbol{Q}_2 g(\mathcal{W})\right] \\
&= \boldsymbol{G}_1\boldsymbol{H}^{\top}\boldsymbol{G}_2\lim_{n\to\infty}\mathrm{E}\left[\mathrm{E}\left[\boldsymbol{Q}_1^{\top}\boldsymbol{Q}_2 g(\mathcal{W})|\boldsymbol{Q}_2\right]\right] \\
&= \boldsymbol{G}_1\boldsymbol{H}^{\top}\boldsymbol{G}_2\lim_{n\to\infty}\mathrm{E}\left[\boldsymbol{Q}_2 g(\mathcal{W})\,\mathrm{E}\left[\boldsymbol{Q}_1^{\top}|\boldsymbol{Q}_2\right]\right] \\
&= \boldsymbol{G}_1\boldsymbol{H}^{\top}\boldsymbol{G}_2\lim_{n\to\infty}\mathrm{E}\left[\boldsymbol{Q}_2 g(\mathcal{W})(\boldsymbol{Q}_2-\boldsymbol{\delta})^{\top}\boldsymbol{G}_2\boldsymbol{H}\boldsymbol{G}_1\right] \\
&= \boldsymbol{G}_1\boldsymbol{H}^{\top}\boldsymbol{G}_2\lim_{n\to\infty}\mathrm{E}\left[\boldsymbol{Q}_2\boldsymbol{Q}_2^{\top} g(\mathcal{W})\right]\boldsymbol{G}_2\boldsymbol{H}\boldsymbol{G}_1 \\
&\quad - \boldsymbol{G}_1\boldsymbol{H}^{\top}\boldsymbol{G}_2\lim_{n\to\infty}\mathrm{E}\left[\boldsymbol{Q}_2 g(\mathcal{W})\right]\boldsymbol{\delta}^{\top}\boldsymbol{G}_2\boldsymbol{H}\boldsymbol{G}_1 \\
&= \boldsymbol{G}_1\boldsymbol{H}^{\top}\boldsymbol{G}_2\boldsymbol{H}\boldsymbol{G}_1\,\mathrm{E}\left[g(\chi^2_{q+2}(\Delta))\right] \\
&\quad + \boldsymbol{\delta}\boldsymbol{\delta}^{\top}\left\{\mathrm{E}\left[g(\chi^2_{q+4}(\Delta))\right] - \mathrm{E}\left[g(\chi^2_{q+2}(\Delta))\right]\right\}.
\end{aligned}$$

Let $\boldsymbol{Z}=\boldsymbol{G}_2^{\frac{1}{2}}\boldsymbol{Q}_2$ once more. Therefore, the last line is then resulted by Lemma 4.2. Also, the third term of Eq. (18) is equal to

$$\begin{aligned}\lim_{n\to\infty}\mathrm{E}\left[\boldsymbol{Q}_4\boldsymbol{Q}_4^\top g^2(\mathcal{W})\right] &= \boldsymbol{G}_1\boldsymbol{H}^\top\boldsymbol{G}_2\lim_{n\to\infty}\mathrm{E}\left[\boldsymbol{Q}_2\boldsymbol{Q}_2^\top g^2(\mathcal{W})\right]\boldsymbol{G}_2\boldsymbol{H}\boldsymbol{G}_1\\ &= \boldsymbol{G}_1\boldsymbol{H}^\top\boldsymbol{G}_2\boldsymbol{H}\boldsymbol{G}_1\,\mathrm{E}\left[g^2(\chi^2_{q+2}(\Delta))\right]+\boldsymbol{\delta}\boldsymbol{\delta}^\top\,\mathrm{E}\left[g^2(\chi^2_{q+4}(\Delta))\right],\end{aligned}$$

For that reason, Eq. (18) is

$$\begin{aligned}\boldsymbol{\Gamma} &= \boldsymbol{G}_1-2\boldsymbol{G}_1\boldsymbol{H}^\top\boldsymbol{G}_2\boldsymbol{H}\boldsymbol{G}_1\,\mathrm{E}\left[g(\chi^2_{q+2}(\Delta))\right]\\ &\quad -2\boldsymbol{\delta}\boldsymbol{\delta}^\top\left\{\mathrm{E}\left[g(\chi^2_{q+4}(\Delta))\right]-\mathrm{E}\left[g(\chi^2_{q+2}(\Delta))\right]\right\}\\ &\quad +\boldsymbol{G}_1\boldsymbol{H}^\top\boldsymbol{G}_2\boldsymbol{H}\boldsymbol{G}_1\,\mathrm{E}\left[g^2(\chi^2_{q+2}(\Delta))\right]+\boldsymbol{\delta}\boldsymbol{\delta}^\top\,\mathrm{E}\left[g^2(\chi^2_{q+4}(\Delta))\right]\\ &= \boldsymbol{G}_1-\boldsymbol{G}_1\boldsymbol{H}^\top\boldsymbol{G}_2\boldsymbol{H}\boldsymbol{G}_1\left\{2\,\mathrm{E}\left[g(\chi^2_{q+2}(\Delta))\right]-\mathrm{E}\left[g^2(\chi^2_{q+2}(\Delta))\right]\right\}\\ &\quad -\boldsymbol{\delta}\boldsymbol{\delta}^\top\left\{2\,\mathrm{E}\left[g(\chi^2_{q+4}(\Delta))\right]-2\,\mathrm{E}\left[g(\chi^2_{q+2}(\Delta))\right]-\mathrm{E}\left[g^2(\chi^2_{q+4}(\Delta))\right]\right\}.\end{aligned}$$

So,

$$\begin{aligned}\mathrm{ADR}(\hat{\boldsymbol{\beta}}^{\text{shrinkage}}) &= \mathrm{tr}(\boldsymbol{W}\boldsymbol{G}_1)-\mathrm{tr}(\boldsymbol{W}\boldsymbol{G}^\star)\left\{2\,\mathrm{E}\left[g(\chi^2_{q+2}(\Delta))\right]-\mathrm{E}\left[g^2(\chi^2_{q+2}(\Delta))\right]\right\}\\ &\quad -\boldsymbol{\delta}^\top\boldsymbol{W}\boldsymbol{\delta}\left\{2\,\mathrm{E}\left[g(\chi^2_{q+4}(\Delta))\right]-2\,\mathrm{E}\left[g(\chi^2_{q+2}(\Delta))\right]-\mathrm{E}\left[g^2(\chi^2_{q+4}(\Delta))\right]\right\}.\end{aligned}$$

Using the definition of shrinkage estimators, the proof is completed. □

The ADR of the FGLS estimator does not depend on the Δ while the restricted does. When Δ equals 0,

$$\mathrm{ADR}(\hat{\boldsymbol{\beta}})-\mathrm{ADR}(\tilde{\boldsymbol{\beta}})=-\,\mathrm{tr}(\boldsymbol{W}\boldsymbol{G}^\star)<0.$$

Thus, the restricted estimator strictly dominates the FGLS estimator when the null hypothesis is true. If Δ moves away from zero, then

$$\mathrm{ADR}(\hat{\boldsymbol{\beta}})-\mathrm{ADR}(\tilde{\boldsymbol{\beta}})=-\,\mathrm{tr}(\boldsymbol{W}\boldsymbol{G}^\star)+\boldsymbol{\delta}^\top\boldsymbol{W}\boldsymbol{\delta}.$$

The restricted estimator outperforms the FGLS estimator for $\boldsymbol{\delta}^\top\boldsymbol{W}\boldsymbol{\delta}<\mathrm{tr}(\boldsymbol{W}\boldsymbol{G}^\star)$. Consequently, the performance of the restricted estimator will strongly depend on the reliability of the null hypothesis.

For $\alpha\in(0,1)$ and $\Delta>0$, we know that

$$H_{q+4}\left(\chi^2_{q,\alpha};\Delta\right)\le H_{q+2}\left(\chi^2_{q,\alpha};\Delta\right)\le H_{q+2}\left(\chi^2_{q,\alpha};0\right)=1-\alpha.$$

As mentioned before, as $\Delta\to\infty$, $H_{q+4}\left(\chi^2_{q,\alpha};\Delta\right)\to 0$. When Δ is set to 0, $H_{q+2}\left(\chi^2_{q,\alpha};\Delta\right)$ and $g_1(\alpha;\Delta)$ approach to 0 and the ADR of the preliminary test estimator approaches that of the FGLS estimator. The ADR function of the preliminary test is smaller than that of the FGLS estimator; when Δ is near zero, it increases, crosses the ADR function of the FGLS estimator, reaches a maximum point, then decreases monotonically to the ADR of the FGLS estimator. The preliminary test estimator is superior to the FGLS estimator if

$$\boldsymbol{\delta}^\top\boldsymbol{W}\boldsymbol{\delta}<\frac{\mathrm{tr}\left(\boldsymbol{W}\boldsymbol{G}^\star\right)H_{q+2}\left(\chi^2_{q,\alpha};\Delta\right)}{g_1(\alpha;\Delta)}.$$

Note that the ADR function of the restricted estimator is unbounded while its preliminary test estimator controls the risk as described.

Like the preliminary test, the Stein-type estimator enjoys the asymptotic property, i.e. as $\Delta\to\infty$, its risk converges to that of the FGLS estimator. When Δ is near 0, both the restricted and preliminary test estimators dominate the Stein-type estimator. However, in the rest of parameter space, they behave worse than the Stein-type shrinkage FGLS estimator. The ADR functions of the FGLS and Stein-type estimators reveal that $\hat{\boldsymbol{\beta}}^{\mathrm{S}}$ will dominate $\tilde{\boldsymbol{\beta}}$ if

$$\boldsymbol{\delta}^\top\boldsymbol{W}\boldsymbol{\delta}>\frac{\mathrm{tr}(\boldsymbol{W}\boldsymbol{G}^\star)g_2(\Delta)}{(q+2)\,\mathrm{E}\left[\chi^{-4}_{q+4}(\Delta)\right]}. \tag{19}$$

Finally, we compare the ADR functions of both the Stein-type and positive-rule estimators. It is concluded that $\mathrm{ADR}(\hat{\boldsymbol{\beta}}^{\mathrm{S}+}) \leq \mathrm{ADR}(\hat{\boldsymbol{\beta}}^{\mathrm{S}})$ for all Δ. For that reason, the positive-rule Stein estimator dominates the FGLS estimator if the condition (19) satisfied.

5. Monte Carlo Simulation Study

A system regression model is considered as Eq. (1). The matrix $\boldsymbol{X}_k$, $k = 1, \ldots, M$ in this model is made up of p variables generated by a standard p-variate normal model.

For $k = 1, \ldots, M$, $\boldsymbol{\beta}_k^0 = (\mathbf{1}_k, \mathbf{0}_k, \mathbf{1}_k, \mathbf{0}_k, \ldots, \mathbf{1}_k, \mathbf{0}_k)^\top$ where $\mathbf{1}_k = (1, \ldots, 1)^\top$, p_{1k}-vector and $\mathbf{0}_{2k} = (0, 0, \ldots, 0)^\top$, p_{2k}-vector, where $p_{1k} + p_{2k} = p_i$. Also, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, \ldots, \boldsymbol{\varepsilon}_M^\top)^\top \sim \mathcal{N}_M(\mathbf{0}, \boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes I_n)$, where the error covariance matrix $\boldsymbol{\Sigma}$ elements are $\Sigma_{ij} = \rho^{|i-j|}$, $i, j = 1, \ldots, M$. In this study, $\rho \in \{0, 0.2, 0.9\}$.

We are interested in testing the hypothesis $\mathcal{H}_o : \boldsymbol{\beta}_{j2} = \mathbf{0}$ for $j = 1, \ldots, M$. For that reason, the regression coefficients are partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_{11}^\top, \boldsymbol{\beta}_{12}^\top, \boldsymbol{\beta}_{21}^\top, \boldsymbol{\beta}_{22}^\top, \ldots, \boldsymbol{\beta}_{M1}^\top, \boldsymbol{\beta}_{M2}^\top)^\top$ where $\boldsymbol{\beta}_{i1} = \mathbf{1}$ is a fixed p_{i1} vector and $\boldsymbol{\beta}_{i2} = (0, \ldots, 0, \Delta)^\top$ is a p_{2i} vector for different Δ. The degree of deviation from the null hypothesis is obtained by $\Delta^* = \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\|^2$. They are used to generate the response variable $\boldsymbol{Y} = (y_1, \ldots, y_M)^\top$ for some values in $[0, 10]$ of Δ^*.

When $\Delta^* = \mathbf{0}$, the submodel is the true model under the null hypothesis, while $\Delta^* > 0$ means a departure from the hypothesized model. Data have been generated for both $\Delta^* = 0$ and $\Delta^* > 0$.

The risk performance using the relative mean squared error (RMSE) criterion is measured. The RMSE of an estimator $\hat{\boldsymbol{\beta}}_1^*$ is defined as

$$\mathrm{RMSE}(\hat{\boldsymbol{\beta}}_1^*; \tilde{\boldsymbol{\beta}}_1) = \frac{\mathrm{MSE}(\tilde{\boldsymbol{\beta}}_1)}{\mathrm{MSE}(\hat{\boldsymbol{\beta}}_1^*)}, \tag{20}$$

where $\hat{\boldsymbol{\beta}}_1^*$ is one of the estimators listed and $\tilde{\boldsymbol{\beta}}$ is the FGLS estimator. The presence of RMSE values greater than 1 indicates that the estimator $\hat{\boldsymbol{\beta}}_1^*$ outperforms $\tilde{\boldsymbol{\beta}}_1$. Several simulation datasets were generated to investigate the behavior of the estimators in the proposed model. The results are the same. Thus they are reported only for particular values in Figure 1.

Based on Figure 1, we can find that

(1) The restricted estimator has the best performance at $\Delta^* = 0$ when the null hypothesis is true.

(2) When Δ^* becomes larger, the RMSE of the restricted estimator decreases and quickly behaves worse than that of the FGLS estimator.

(3) With an increase in Δ^*, all other estimators perform better than the FGLS estimator.

(4) When Δ^* obtains larger values, the preliminary test, Stein-type, and positive-rule Stein-type estimators behave the same as the FGLS estimator.

(5) The preliminary test estimator has a good performance near the null hypothesis; although this is true, it depends upon how close $\boldsymbol{\beta}_{i2}$ is to zero vectors for $i = 1, \ldots, M$.

(6) Both shrinkage estimators have better performance concerning the FGLS estimator for all values of Δ^*.

The LASSO is used as a variable selection method in this simulation study. It produced a submodel and was employed to construct the preliminary test and shrinkage estimators. Figure 2 is a report of the performance of this simulation, which is consistent with the results of Figure 1. Consequently, we recommend using the proposed estimators when (1) there is a high correlation between covariates and (2) when the number of nuisance parameters is large.

5.1. Application

In economics, the employees' willingness to pay is important. Employers frequently want to attract good employees by offering attractive compensation packages. On the other hand, compensation extends beyond an employee's hourly or salary wages. It also includes fringe benefits, which are additional small business employee perks. Fringe benefits are the additional benefits offered to an employee above the stated salary to perform a specific service. Some fringe benefits such as social security and health insurance are required by law, while the employer voluntarily provides others. Free breakfast and lunch, gym membership, employee stock options, transportation benefits, retirement planning services, childcare, and education assistance are all examples of optional fringe benefits. So, one of valuable questions is "Do the employees prefer wages or fringes?". The answer is fundamental to future planning. Invariably, a trade-off between wage and fringes is an exciting topic for economics; using SUR models simultaneously leads to an investigation of both. Also, Injecting additional information into the model improves the accuracy.

In this section, the dataset "Fringe†", wages and fringe benefits, are used to estimate a two-equation regression system for hourly wage and hourly benefits.

†Available at http://fmwww.bc.edu/ec-p/data/wooldridge/fringe.dta

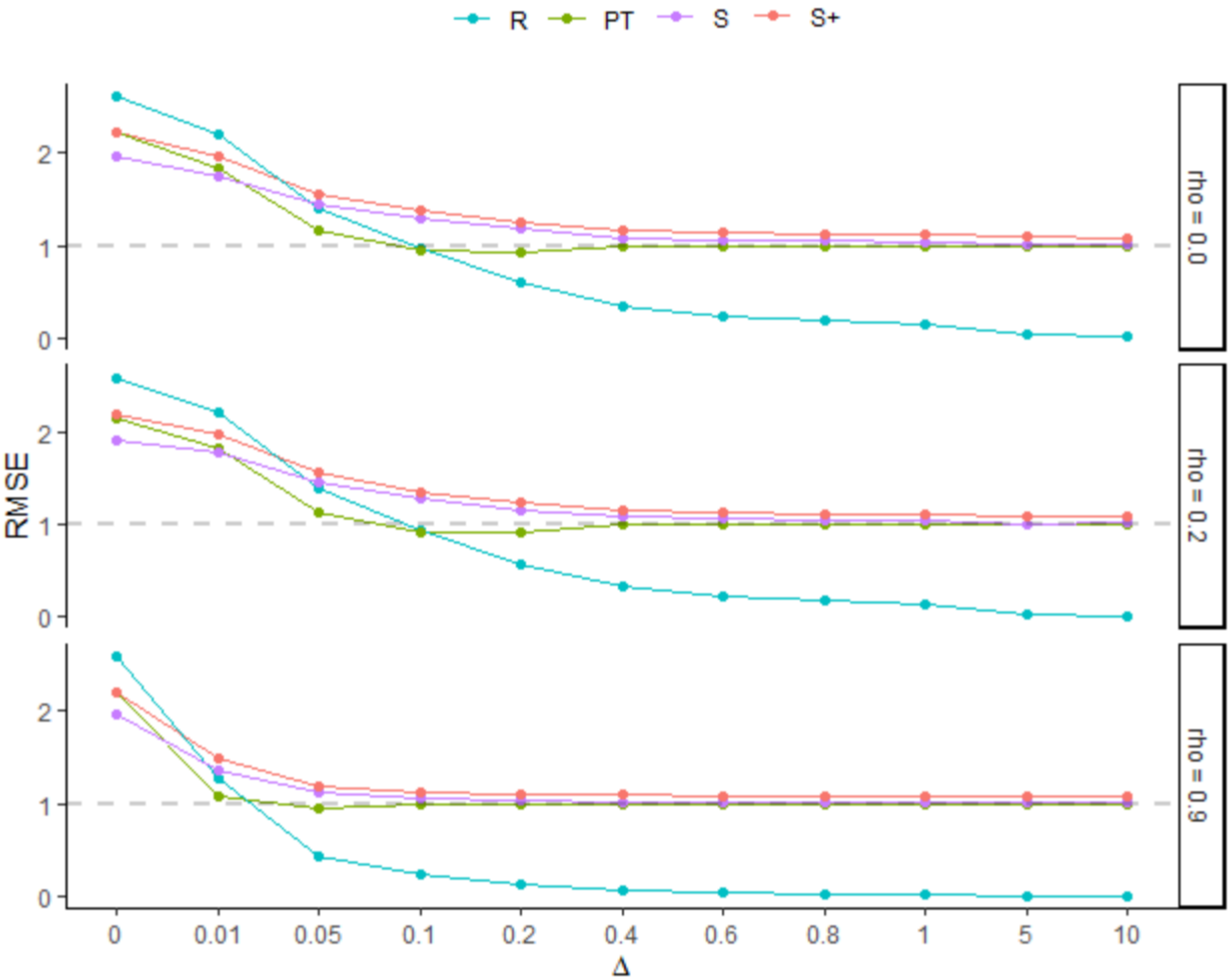

Figure 1. The RMSE of the estimators with respect to the FGLS estimator for $n = 100$, $M = 3$, $p_{11} = p_{21} = p_{31} = 2$, $p_{12} = p_{22} = p_{32} = 3$, and $\alpha = 0.05$ for different ρ.

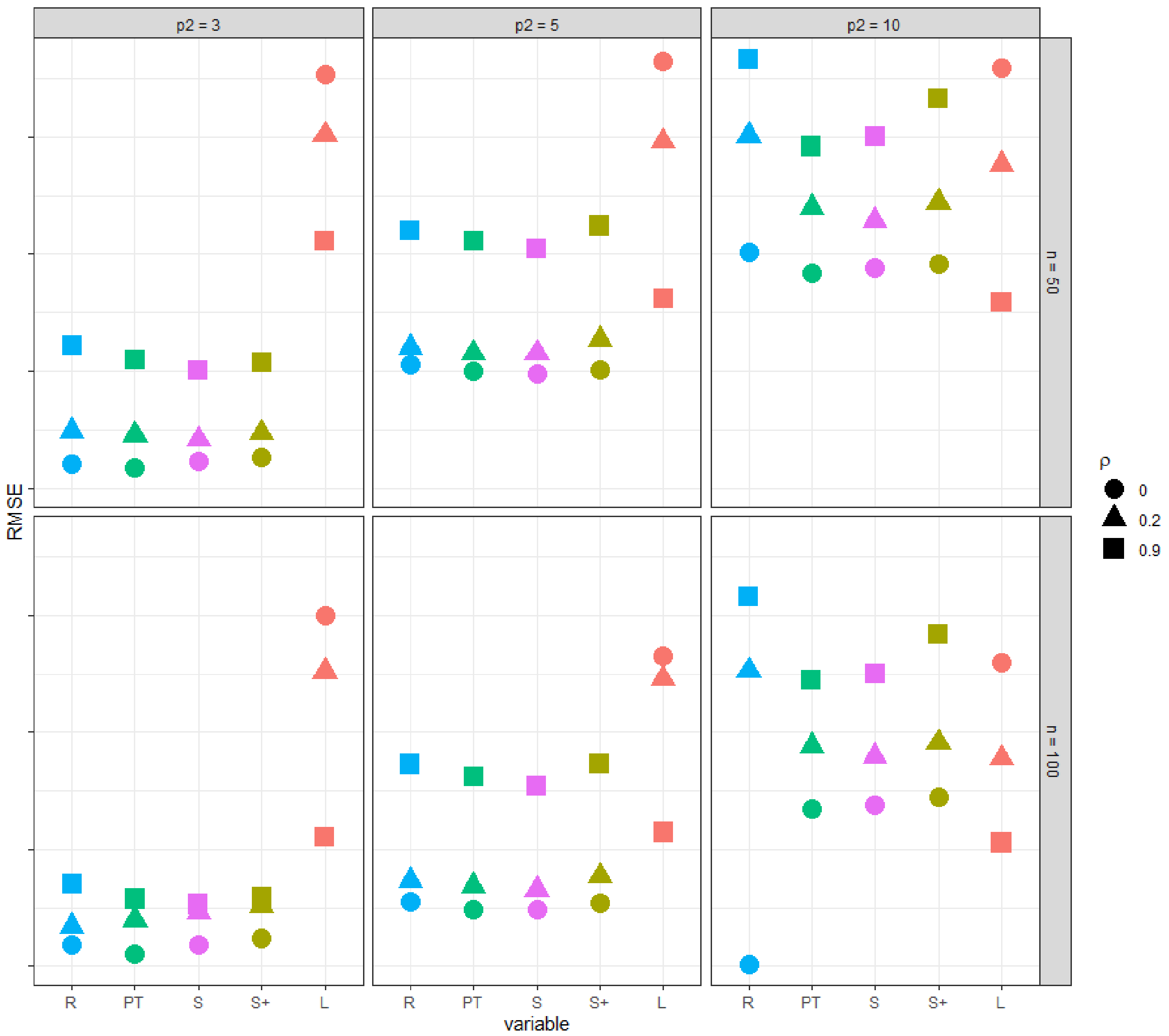

Figure 2. The RMSE with respect to FGLS estimator for $p_1 = 5$, $\Delta^* = 0$ and $\alpha = 0.05$.

There are 616 workers in the dataset. There are two response variables, here: hearn, hourly earning ($), and hrbens, hourly benefits ($). The explanatory variables for both of the equations are: educ, years schooling; exper, years work experience; tenure, years with current employer; union, = 1 if a union member; south, = 1 if live in south; nrthcen, = 1 if live in north central; nrtheast, = 1 if live in northeast; married, = 1 if married; white, = 1 if white; male, = 1 if male, age, age in years; office, = 1 if an office worker.

The AIC, BIC, and LASSO approaches have selected the important predictors. Selected variables by each method are presented in Table 2. Thus, there are two models: a full model with all covariates and three suspecting sub-models based on the AIC, BIC, and LASSO methods, which exclude the nuisance variables.

Table 2. Selected variables of the "Fringe" dataset by the AIC, BIC and LASSO approaches.

AIC	
educ ✓	educ ✓
exper✓	exper
tenure	tenure ✓
union✓	union✓
south	south
nrtheast✓	nrtheast
nrthcen	nrthcen
married	married ✓
white✓	white
male ✓	male✓
age ✓	age
office✓	office✓

BIC	
educ ✓	educ✓
exper ✓	exper
tenure	tenure ✓
union ✓	union ✓
south	south
nrtheast	nrtheast
nrthcen	nrthcen
married	married
white ✓	white
male ✓	male ✓
age	age
office ✓	office ✓

LASSO	
educ ✓	educ ✓
exper✓	exper
tenure	tenure✓
union✓	union ✓
south	south
nrtheast✓	nrtheast
nrthcen	nrthcen
married✓	married
white ✓	white
male ✓	male✓
age	age
office✓	office✓

To compare estimators, the K-fold cross validation method is employed. In this method, K different equal and random subsets of the dataset are considered. One of the subsets, $\{(\boldsymbol{X}^{\text{test}}, \boldsymbol{y}^{\text{test}})\}$, a test set is left and other subsets, training sets, are used to fit the model. The result estimator is called $\hat{\boldsymbol{\beta}}^{\text{train}}$. Using this estimator, the responses of the test dataset can be estimated. The prediction error (PE) is given by

$$\text{PE}^k = \left\| \boldsymbol{y}_k^{\text{test}} - \hat{\boldsymbol{y}}_k^{\text{test}} \right\|^2; \qquad k = 1, \ldots, K,$$

where $\hat{\boldsymbol{y}}_k^{\text{test}} = \boldsymbol{X}_k^{\text{test}} \hat{\boldsymbol{\beta}}_k^{\text{train}}$. The process is repeated for all K subsets, and the prediction errors are combined via their mean (say PE_{CV}) is obtained by

$$\text{PE}_{\text{CV}} = \frac{1}{K} \sum_{i=1}^{k} \text{PE}^k.$$

It varies across runs for different values of K because the cross validation method is a random procedure. So, the average of PE_{CV} (APE) is considered as a criterion for comparison. The below formula shows how APE is calculated.

$$\text{APE} = \frac{1}{N} \sum_{j=1}^{N} \text{PE}_{\text{CV}}^j,$$

where PE_{CV}^j is the value obtained in jth iteration and N is the number of repetitions.

The performance of the arbitrary estimator, $\hat{\boldsymbol{\beta}}_1^\star$ with respect to the full model estimator $\tilde{\boldsymbol{\beta}}_1$ is obtained by the Efficiency (Eff) formula defined as

$$\text{Eff}(\hat{\boldsymbol{\beta}}_1^\star; \tilde{\boldsymbol{\beta}}_1) = \frac{\text{APE}(\tilde{\boldsymbol{\beta}}_1)}{\text{APE}(\hat{\boldsymbol{\beta}}_1^\star)}.$$

If the value of Eff is greater than 1, then $\hat{\boldsymbol{\beta}}_1^\star$ will perform better than $\tilde{\boldsymbol{\beta}}$.

The relative efficiency based on 10-fold cross-validation of size 1000 is reported in Table 2. We determine that if we use the LASSO procedure only as a selection operator, the restricted estimator performs better than other methods since it can select the active variables included in the submodel, which are indeed/nearly indeed important. It is preferable to use BIC to select data to obtain the best shrinkage estimator. Moreover, the shrinkage estimator based on the LASSO procedure is still performing better than the full model estimator.

Table 3. RAPEs based on 10-fold cross validation

Selection Criterion	$\hat{\boldsymbol{\beta}}_1$	$\hat{\boldsymbol{\beta}}_1^{\mathrm{PT}}$	$\hat{\boldsymbol{\beta}}_1^{\mathrm{S+}}$	$\hat{\boldsymbol{\beta}}_1^{\mathrm{LASSO}}$
AIC	1.0227	1.0161	1.0176	0.6042
BIC	1.3908	1.3211	1.3758	0.8191
LASSO	1.6093	1.0000	1.2088	0.9495

6. Conclusion

This paper proposes preliminary test and shrinkage estimators in a SUR model regression based on the feasible generalized least squares estimator. In comparison to FGLS estimation, our suggested estimators show better performance. The simulation results especially recommend using the positive rule shrinkage feasible generalized least squares estimator in two cases: (i) there exists a high correlation between covariates and, (ii) in a sparse model, the number of nuisance parameters is large. The proposed estimator behaves better than the LASSO estimator. Although selecting important variables by the LASSO method yields less error in prediction, the positive-rule Stein-type shrinkage FGLS estimator obtained by the BIC approach predicts better than that achieved by the LASSO. The BIC method is consistent. However, when the number of parameters gets large, it is less efficient than the LASSO. For that reason, and because of the better performance of the positive-rule Stein-type FGLS estimator, we suggest using this estimator in a SUR model. Other penalized estimators, such as Enet [45] estimator, can be used instead of LASSO to select variables in future research.

Acknowledgement

The current form of the manuscript is a result of the editor's and referees' insightful comments during the reading process. The authors gratefully acknowledge the reviewer's helpful comments and suggestions, which have improved this paper.

M. Norouzirad and F. J. Marques wish to acknowledge funding provided by National Funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the projects UIDB/00297/2020 and UIDP/00297/2020 (Center for Mathematics and Applications).

REFERENCES

1. M. A. Alkhamisi, and G. Shukur, Developing ridge parameters for SUR model, Communications in Statistics–Theory and Methods, vol. 37, pp. 544–564, 2008.
2. T. Amemiya, Advanced Econometrics, Cambridge, Massachusetts: Harvard University Press, 1985.
3. M. Arashi, S. Khan, S. M. M. Tabatabaey, and H. Soleimani, Shrinkage estimation under multivariate elliptic models, Communications in Statistics – Theory and Methods, vol. 42, no. 11, pp. 2084–2103, 2013.
4. M. Arashi, B. G. Kibria, M. Norouzirad, and S. Nadarajah, Improved preliminary test and stein-rule liu estimators for the ill-conditioned elliptical linear regression model, Journal of Multivariate Analysis, vol. 126, pp. 53–74, 2014.
5. M. Arashi, and Roozbeh, M., Shrinkage estimation in system regression model. Computational Statistic, vol. 30, pp. 359–376, 2015.
6. M. Arashi and M. Roozbeh, Some improved estimation strategies in high-dimensional semiparametric regression models with application to riboflavin production data, Statistical Papers, vol. 60, pp. 667–686, 2019.
7. M. Arashi, M. Roozbeh, M., and H. A. Niroumand, A note on stein-type shrinkage estimator in partial linear models. Statistics, vol. 46, no. 5, pp. 673–685, 2012.
8. M. Arashi, and S. M. M. Tabatabaey, A note on classical stein-type estimators in elliptically contoured models, .Journal of Statistical Planning and Inference, vol. 140, no. 5, pp. 1206–1213, 2010.
9. D. G. Fiebig, Seemingly unrelated regression. In H. Badi (Ed.) A Companion to Theoretical Econometrics, Baltagi, Oxford: Blackwell.
10. W. H. Green, Econometric Analysis (5th ed.). New Jersey: Pearson/Prentice Hall, 2003.
11. W. James, and C. Stein, Estimation with quadratic loss, in:. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 361 – 380, 1961.

12. G. G. Judge and M. E. Bock, The Statistical Implication of Pretest and Stein Rule Estimators in Econometrics, Amsterdam: North-Holland, 1978.
13. M. Kashani, M. R. Rabie, M. Arashi, An integrated shrinkage strategy for improving efficiency in fuzzy regression modeling, Soft Computing, vol. 25, pp. 8095-–8107, 2021.
14. J. Kleyn, M., Arashi, A., Bekker, and S. Millard, Preliminary testing of the cobb–douglas production function and related inferential issues, Communications in Statistics – Simulation and Computation, vol. 46, no. 1, pp. 469–488, 2017.
15. A. Mahmoudi, R. A., Belaghi, and S. Mandal, A comparison of preliminary test, stein-type and penalty estimators in gamma regression model, Journal of Statistical Computation and Simulation, vol. 90, no. 17, pp. 3051–3079, 2022.
16. S. Mandal, R. Arabi Belaghi, A. Mahmoudi, and M. Aminnejad, Stein-type shrinkage estimators in gamma regression model with application to prostate cancer data, Statistics in Medicine, vol. 38, no. 22, pp. 4310–4322, 2019.
17. A. Mehrabani, and A. Ullah, Improved average estimation in seemingly unrelated regression, Econimetrics, vol. 8, no. 2, pp. 15, 2020.
18. M. NooriAsl, H. Bevrani, and R. Arabi Belaghi, Penalized and ridge-type shrinkage estimators in Poisson regression model, Communications in Statistics – Simulation and Computation, vol. 5, no. 7, pp. 4039–4056, 2022.
19. M. Norouzirad, and M. Arashi, Preliminary test and stein-type shrinkage lasso-based estimators, SORT-Statistics and Operations Research Transactions, vol. 42, no. 1, pp. 45–58, 2018.
20. M. Norouzirad, and M. Arashi, Preliminary test and stein-type shrinkage ridge estimators in robust regression, Statistical Papers, vol. 60, pp. 1849–1882, 2019.
21. M. Norouzirad, M. Arashi, and S. E. Ahmed, Improved robust ridge M-estimation, Journal of Statistical Computation and Simulation, vol. 87, no. 18, pp. 3469–3490.
22. M. Norouzirad, M. Arashi, and M. Roozbeh, Differenced-based double shrinking in partial linear models, Journal of Computational Statistics and Modeling, vol. 1, no. 1, pp. 21–32, 2021.
23. M. Norouzirad, S. Hossain, and M. Arashi, Shrinkage and penalized estimators in weighted least absolute deviations regression models, Journal of Statistical Computation and Simulation, vol. 88, no. 8, pp. 1557–1575, 2018.
24. M. Roozbeh, Shrinkage ridge estimators in semiparametric regression models, Journal of Multivariate Analysis, vol. 136, pp. 56–74, 2015.
25. A. Safariyan, M., Arashi, R. Arabi Belaghi, Improved estimators for stress-strength reliability using record ranked set sampling scheme, Communications in Statistics – Simulation and Computation, vol. 48, no. 9, pp. 2708–2726, 2019.
26. A. Safariyan, M., Arashi, R. Arabi Belaghi, Improved point and interval estimation of the stress–strength reliability based on ranked set sampling Statistics, vol. 53, no. 1, pp. 101–125.
27. A. K. M. E. Saleh, M. Arashi, B. M. G. Kibria, Theory of Ridge Regression Estimation with Applications, New Jersey: Wiley, 2019.
28. A. K. M. E. Saleh, M. Arashi, R. A. Saleh, and M. Norouzirad, Rank-Based Methods for Shrinkage and Selection: With Application to Machine Learning, John Wiley and Sons, 2022.
29. A. K. M. E. Saleh, M. Arashi, and S. M. M. Tabatabaey, Statistical Inference for Models with Multivariate t-Distributed Errors, New Jersey: Wiley, 2014.
30. A. K. M. E. Saleh, Theory of preliminary test and stein-type estimation with applications, New York: John Wiley & Sons, 2006.
31. A. K. M. E. Saleh, R. Navrátil, and M. Norouzirad, Rank theory approach to ridge, lasso, preliminary test and stein-type estimators: A comparative study, Canadian Journal of Statistics, vol. 46, no. 4, pp. 690–704, 2018.
32. A. K. M. E. Saleh, and M. Norouzirad, On shrinkage estimation: Non-orthogonal case, Statistics, Optimization and Information Computing, vol. 6, no.3, pp. 427–451, 2018.
33. S. L. Sclove, Improved estimators for coefficients in linear regression, Journal of American Statistical Association, vol. 63, pp. 596- 606, 1968.
34. D. Sengupta, and S. R. Jammalamadaka, Linear Models: An Integrated Approach, World Scientific Publishing Company, 2003.
35. Shalabh Regression analysis,. Lectures Notes for National Digital Library of India. http://www.ndl.gov.in/document/YkxlRXFvZXJrTDBkVzVVZi9ESjl6L01uL3BEcDB2QzV1NHRUcGNLUXF0SVp6L3ZSUUVudUlNYitlZ1lRU3NKUQ, 2020
36. V. K. Srivastava, The efficiency of an improved method of estimating seemingly unrelated regression equations, Journal of Econometrics, vol. 1, pp. 341–350, 1973.
37. V. K. Srivastava, and T. D. Dwivedi, Estimation of seemingly unrelated regression equations: A brief survey, Journal of Econometrics, vol. 10, pp. 15–32, 1979.
38. V. K. Srivastava, and D. E. A. Giles, Seemingly Unrelated Regression Models: Estimation and Inference, Marcel Dekker Inc.: New York, 1987.
39. V. K. Srivastava, and A. T. K. Wan, Seperate versus system methods of stein-rule estimation in seemingly unrelated regression models, Communications in Statistics – Theory and Methods, vol. 31, no. 11, pp. 2077–2099, 2002.
40. R. Tibshirani, Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society (Series B), vol. 58, pp. 267–288, 1996.
41. B. Yuzbasi, and S. E. Ahmed, Ridge type shrinkage estimation of seemingly unrelated regressions and analytics of economic and financial data from “fragile five” countries, Journal of Risk and Financial Management, vol. 13, pp. 131, 2020.
42. Z. Zandi, H. Bevrani, and R. Arabi Belaghi, Estimation of fixed parameters in negative binomial mixed model using shrinkage estimators, Journal of Computational Statistics and Modeling, vol. 1, no. 2, pp. 99–124, 2021.
43. A. Zellner, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, Journal of the American Statistical Association, vol. 57, no. 298, pp. 348 – 368, 1962.

44. A. Zellner, and W. Vandaele, Bayes-stein estimators for k-means, regression and simultaneous equation models, In S. E. Fienberg, and A. Zellner (Eds.) Studies in Bayesian Econometrics and Statistics, (pp. 627–653). Amsterdam: North-Holland Publishing Company.
45. H. Zou, H., and T. Hastie, Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society. Series B (Statistical Methodology), vol. 67, no. 2, pp. 301–320, 2005.