Journal of Statistical Research *https://doi.org/10.3329/jsr.v59i1.83690*
2025, Vol. 59, No. 1, pp. 131-143 *ISSN 0256 - 422 X*

THEORETICAL DEVELOPMENT OF SHRINKAGE LEARNERS IN THE SEEMINGLY UNRELATED SEMIPARAMETRIC MODEL

MOHAMMAD ARASHI*

Department of Statistics, Faculty of Mathematical Sciences
Ferdowsi University of Mashhad, Iran

Email: arashi@um.ac.ir

MAHDI ROOZBEH

Department of Statistics, Faculty of Mathematics, Statistics and Computer Sciences
Semnan University, Iran

Email: mahdi.roozbeh@semnan.ac.ir

MORTEZA AMINI

Department of Statistics, School of Mathematics, Statistics, and Computer Science
College of Science, University of Tehran, Tehran, Iran

Email: morteza.amini@ut.ac.ir

SUMMARY

While the existing literature includes substantial numerical investigations into various shrinkage ridge and Liu estimators, it often lacks a cohesive approach to their construction. This gap signals a need for a unified construction methodology that can provide a clearer framework for understanding and applying Liu estimators in practice. By establishing such a methodology, we aim to simplify the utilization of these estimators and promote their adoption in various statistical applications. This paper will discuss the theoretical underpinnings of shrinkage learners with focus on the seemingly unrelated semiparametric regression model. Through this construction analysis, we ultimately aim to enhance the ongoing discourse in the field of shrinkage learners, offering valuable insights that support researchers and practitioners in choosing suitable techniques for their specific data challenges.

Keywords and phrases: Liu estimator; Multicollinearity; Partially linear model; Preliminary test estimator; Seemingly unrelated regression; Semiparametric; Shrinkage estimator; Stein-type estimator

* Corresponding author

1 Introduction

The seemingly unrelated regression (SUR) model introduced by Zellner (1962) consists of multiple individual relationships interconnected by correlated disturbances. These models have numerous applications. For instance, demand functions for several families (or household categories) may be evaluated for a specific commodity. The dependency among the equation disruptions may stem from multiple origins, including synchronized disturbances to household income. Conversely, one may simulate a household's demand for several goods; however, the addition constraints impose limitations on the parameters of the distinct equations in this scenario. Conversely, equations that describe phenomena across various cities, states, countries, enterprises, or industries have significant relevance, as these entities are prone to being affected by spillovers from economy-wide or global shocks.

This research advances the SUR system into a more adaptable framework, namely the seemingly unrelated semiparametric model (SUS), to enhance the modeling efficiency of a data set. There are situations where the data structure cannot be fully represented by either linear or non-parametric models, since both linear and non-linear relationships may exist between the response and explanatory variables in the system of equations. The methodologies, including this model, provide more flexible response prediction by categorizing predictors into linear and non-linear categories employing robust instruments like the added variable plot.In fact, this model extends an applicable semiparametric regression framework through a vectorized parameter approach. There are two primary motivations for employing SUS. The first is to improve estimation efficiency by integrating information across multiple equations. The second is to increase the adaptability of the SUR model, which makes it more useful in applications such as political behavior, including voting, biometric problems, allocation models, investment functions for multiple firms, and income or consumption functions for specific population segments or various geographic regions (see Baltagi (1980), Chib and Greenberg (1995), Fiebig (2001), Moon (1999), Moon and Perron (2004), Srivastava and Giles (1987), and Srivastava and Maekawa (1995)). This model is well-suited for analyzing a data system where the relationship between the dependent variable and some explanatory variables is clearly linear and parametric, while the connection with other explanatory variables is uncertain and explicitly non-parametric. Refer to Roozbeh et al. (2012) and Roozbeh and Arashi (2014) for details.

It is widely acknowledged that the ridge regression estimator serves as a robust learner of regression parameters across various regression models. It possesses beneficial characteristics, notably its ability to deliver continuous predictions. However, two significant drawbacks have prompted researchers to explore the Liu estimator as an alternative. First, the ridge estimator exhibits nonlinearity concerning the tuning parameter, which complicates the optimization process. In a study, Kibria (2022) provides an extensive discussion of over one hundred alternative estimators. More critically, the impact of the tuning parameter poses challenges; in situations of severe multicollinearity, the ridge estimator may fall short in addressing the issues associated with an ill-conditioned design matrix. To address these limitations, Liu (2003) introduced the Liu-type penalty for linear regression models, aiming to mitigate the bias inherent in the ridge method and better manage high multicollinearity. The Liu estimator has garnered significant attention in efforts to develop estimators that are resilient to multicollinearity. This is evident in several studies, including Kibria (2003), Arashi et al. (2017), Asar et al. (2017), Arashi et al. (2018), Wu et al. (2018), Qasim et al. (2019), Lukman et al. (2020), Karlsson et al. (2020), Amin et al. (2021), Algamal and Abonazel (2022), Arashi et al. (2022), Al-Momani (2023), Abonazel (2023), Akram et al. (2024), Ghanem et al. (2024), Tanış and Asar (2024), Altukhaes et al. (2024b), Altukhaes et al. (2024a), Genç and Lukman (2025), and Hawa et al. (2025) to mention a few.

In this paper, we contribute only to theoretical construction of Liu-type shrinkage strutegies for learning the parameters of seemingly unrelated semiparametric regression model. This contribution is made to honor and acknowledge ground breaking contributions of late Professor A. K. Md. Ehsanes Saleh, in the field of shrinkage strategies.

We organize the remainder of this paper as follows: Section 2 presents the elliptical seemingly unrelated regression model along with its assumptions. To ensure generality, we focus on partially linear models, where non-linear effects are incorporated into the modeling of the expected response variable. The construction methodology of shrinkage learners is discussed in detail in Section 3, while section 4 contains the proofs of two propositions. Finally, a brief conclusion is provided in Section 5.

2 The Model and Assumptions

In this section, we confine ourselves to the notation of Roozbeh et al. (2012). We define the model and give necessary preliminary results on the least squares theory of estimation in the partially linear SUR models.

Consider a system of M equations expressed as $\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{f}_i(\boldsymbol{t}) + \boldsymbol{\epsilon}_i$, for $i = 1, \ldots, M$, where $\boldsymbol{Y}_i$ represents an $n \times 1$ response vector, $\boldsymbol{X}_i = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$ is an $n \times p_i$ fixed design matrix with full rank p_i, constructed from the explanatory variables $\boldsymbol{x}_j \in \mathbb{R}^{p_i}$, $j = 1, \ldots, n$. The vector $\boldsymbol{\beta}_i$ consists of $p_i \times 1$ regression coefficients, while $\boldsymbol{f}_i(\boldsymbol{t})$ denotes a $n \times 1$ vector of unknown smooth functions, and $\boldsymbol{\epsilon}_i$ is a $n \times 1$ vector of random error terms. It is assumed that $\mathbb{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_j^\top) = v_{ij}\boldsymbol{I}_n$, where v_{ij} indicates the inter-equation covariance for each observation. In this formulation, M is the total number of equations, n denotes the sample size per equation, and p_i refers to the number of parameters in $\boldsymbol{\beta}_i$.

Considering the shrinkage estimators as one of the main parts of machine learners (refer to Saleh et al. (2022)), we develop Stein-type shrinkage estimators under subspace restrictions. To this aim, we aggregate the model parameters, and assume $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top, \ldots, \boldsymbol{Y}_M^\top)^\top$, $\boldsymbol{X} = \mathrm{Diag}(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \ldots, \boldsymbol{\beta}_M^\top)^\top$, $\boldsymbol{f}(\boldsymbol{t}) = \left(\boldsymbol{f}_1^\top(\boldsymbol{t}), \boldsymbol{f}_2^\top(\boldsymbol{t}), \ldots, \boldsymbol{f}_M^\top(\boldsymbol{t})\right)^\top$ and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^\top, \boldsymbol{\epsilon}_2^\top, \ldots, \boldsymbol{\epsilon}_M^\top)^\top$. Then, the system of M equations can be consolidated into the following representation

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{f}(\boldsymbol{t}) + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\boldsymbol{Y}$, $\boldsymbol{f}(\boldsymbol{t})$ and $\boldsymbol{\epsilon}$ are each of dimension $nM \times 1$, $\boldsymbol{X}$ is of dimension $nM \times p$, $p = \sum_{i=1}^{M} p_i$, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters. Furthermore, $\mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0}$, and $\mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) = \boldsymbol{V}$, with

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

as an $nM \times nM$ positive definite symmetric matrix, or simply as $(v_{ij}) \otimes \boldsymbol{I}_n$, where $\otimes$ represents the Kronecker product. This assumption enforces homoscedasticity between the errors and not autocorrelation. However, it is assumed that there is a simultaneous correlation between the matching error terms in different equations.

For the estimation of $\boldsymbol{\beta}$, we use the partial kernel smoothing estimator and thus, using Speckman (1988), we estimate $f_i(\cdot)$ via

$$\hat{f}_i(t_j, \boldsymbol{\beta}_i) = \sum_{k=1}^{n} W_{n_k}(t_j)(y_{ik} - \boldsymbol{x}_{ik}^\top \boldsymbol{\beta}_i),$$

where the positive weight functions $W_{n_k}(\cdot)$ satisfy $\max \sum_{j=1}^{n} W_{n_k}(t_j) = O(1)$, $\max W_{n_k}(t_j) = O(n^{-2/3})$, and $\max \sum_{k=1}^{n} W_{n_k}(z_j) I(|t_i - t_j| > c_n) = O(d_n)$, for some sequences c_n and d_n satisfying $\limsup nc_n^3 < \infty$, and $\limsup nd_n^3 < \infty$.

Now, define

$$\begin{aligned} \boldsymbol{C} &= \widetilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1} \widetilde{\boldsymbol{X}} = \mathrm{Diag}(\boldsymbol{C}_1, \dots, \boldsymbol{C}_M), \quad \boldsymbol{C}_i = \widetilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_{ii}^{-1} \widetilde{\boldsymbol{X}}_i, \\ \widetilde{\boldsymbol{Y}} &= (\widetilde{y}_{11}, \dots, \widetilde{y}_{1n}, \widetilde{y}_{21}, \dots, \widetilde{y}_{2n}, \dots, \widetilde{y}_{M1}, \dots, \widetilde{y}_{Mn})^\top, \\ \widetilde{\boldsymbol{X}} &= (\widetilde{\boldsymbol{x}}_{11}, \dots, \widetilde{\boldsymbol{x}}_{1n}, \widetilde{\boldsymbol{x}}_{21}, \dots, \widetilde{\boldsymbol{x}}_{2n}, \dots, \widetilde{\boldsymbol{x}}_{M1}, \dots, \widetilde{\boldsymbol{x}}_{Mn})^\top, \\ \widetilde{y}_{ij} &= y_{ij} - \sum_{j=1}^{nM} W_{n_j}(t_i) y_{ij}, \\ \widetilde{\boldsymbol{x}}_{ij} &= \boldsymbol{x}_{ij} - \sum_{j=1}^{nM} W_{n_j}(t_i) \boldsymbol{x}_{ij}, \quad i = 1, \dots, M, \ j = 1, \dots, n. \end{aligned}$$

To estimate $\boldsymbol{\beta}$, we minimize the least squares criterion.

$$SS(\boldsymbol{\beta}) = (\widetilde{\boldsymbol{Y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1} (\widetilde{\boldsymbol{Y}} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})$$

to get the weighted least square estimator (WLSE) given by

$$\hat{\boldsymbol{\beta}}^{SUS} = \boldsymbol{C}^{-1} \widetilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1} \widetilde{\boldsymbol{Y}}. \tag{2.2}$$

3 Shrinkage Linear Unified Machine Learners

The properties of the WLSE of $\boldsymbol{\beta}$ are strongly influenced by the characteristics of the information matrix $\boldsymbol{C}$. If the $\boldsymbol{C}$ matrix is poorly conditioned, then the WLSE yields excessively high sampling variances. On the other hand, some of the coefficients might be statistically insignificant with incorrect directions, making reliable statistical inference challenging for the researcher. As a solution, Roozbeh et al. (2012) introduced the ridge-weighted least square estimator (RWLSE). As a shrinkage method, the RWLSE can be viewed as a standard approach to addressing multicollinearity.However, a small ridge parameter is insufficient to address the ill-conditioned design matrix in cases of severe multicollinearity. On the other hand, increasing the ridge parameter excessively introduces greater bias into the estimation. Motivated by Liu (1993) and Akdeniz and Kaciranlar (1995), we find a regularization parameter matrix $\boldsymbol{K}$ such that $\boldsymbol{K}\hat{\boldsymbol{\beta}}^{\mathrm{SUS}}$ is close to $\boldsymbol{\beta}$; thus, learning the magnitude penalization of $(\boldsymbol{K}\hat{\boldsymbol{\beta}}^{\mathrm{SUS}} - \boldsymbol{\beta})$ rather than $\boldsymbol{\beta}$ as in the RWLSE.

Thus, we propose the following linear unified (Liu) weighted least square estimator (LWLSE)

$$\hat{\boldsymbol{\beta}}^{\mathrm{SUS}}(\boldsymbol{K}) = \boldsymbol{T_K}\hat{\boldsymbol{\beta}}^{\mathrm{SUS}}, \tag{3.1}$$

where $\boldsymbol{T_K} = (\boldsymbol{C} + \boldsymbol{I}_p)^{-1}(\boldsymbol{C} + \boldsymbol{K})$, $\boldsymbol{K} = \mathrm{Diag}(\boldsymbol{K}_1, \boldsymbol{K}_2, \ldots, \boldsymbol{K}_M)$, $\boldsymbol{K}_i = k_i\boldsymbol{I}_{p_i}, k_i \geq 0$ are the shrinking parameters for $i = 1, .., M$, and $p = \sum_{i=1}^{M} p_i$.

Now, consider some non-sample prior information derived from theoretical or experimental insights, testable hypotheses, or intentionally imposed constraints aimed at reducing or removing redundancy in the model's description. Such information, also known as uncertain prior information or restrictions, is useful in the estimation task, especially when there exists limited information based on the sample data.

Adopting such cases, some linear non-stochastic constraints are considered for computation of the restricted machine learner. The following result gives the mathematical form of this learner.

Proposition 3.1. Assume the partially linear SUR model (2.1). Also, consider a set of restrictions $\boldsymbol{R}_i\boldsymbol{\beta}_i = \boldsymbol{r}_i$, $i = 1, \ldots, M$, for a given $m_i \times p_i$ matrix $\boldsymbol{R}_i$ with rank $m_i < p_i$ and a given $m_i \times 1$ vector $\boldsymbol{r}_i$. Then, the restricted Liu machine learner is given by

$$\hat{\boldsymbol{\beta}}^{\mathrm{SUS}}(\boldsymbol{K}) - (\boldsymbol{C} + \boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top(\boldsymbol{R}(\boldsymbol{C} + \boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top)^{-1}\left[\boldsymbol{R}\hat{\boldsymbol{\beta}}^{\mathrm{SUS}}(\boldsymbol{K}) - \boldsymbol{r}\right],$$

where $\boldsymbol{R} = \mathrm{Diag}(\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_M)$ and $\boldsymbol{r} = (\boldsymbol{r}_1^\top, \boldsymbol{r}_2^\top, \ldots, \boldsymbol{r}_M^\top)^\top$ are $m \times p$ and $m \times 1$ matrices, respectively, with $m = \sum_{i=1}^{M} m_i$.

For the sketch of proof, refer to the Appendix. The assumption of full-row rank is adopted for convenience and is justifiable by the fact that any consistent linear equation can be converted into an equivalent equation with a coefficient matrix having full-row rank.

In Proposition 3.1, if $\boldsymbol{V}$ is unknown, one can use a feasible estimator for $\boldsymbol{V}$ as

$$\boldsymbol{S} = \frac{1}{nM - (p - m)}(\widetilde{\boldsymbol{Y}} - \widetilde{\boldsymbol{X}}\boldsymbol{b})(\widetilde{\boldsymbol{Y}} - \widetilde{\boldsymbol{X}}\boldsymbol{b})^\top, \quad \boldsymbol{b} = (\widetilde{\boldsymbol{X}}^\top\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}}^\top\widetilde{\boldsymbol{Y}}.$$

Apart from Proposition 3.1 that gives a proper form for the restricted Liu machine learner, similar to the work of Kaciranlar et al. (1999), we define the restricted learner as a weighted form of the WLSE

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{\mathrm{SUS}}(\boldsymbol{K}) = \boldsymbol{T_K}\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{\mathrm{SUS}} = \boldsymbol{T_K}\left[\hat{\boldsymbol{\beta}}^{\mathrm{SUS}} - \boldsymbol{C}^{-1}\boldsymbol{R}^\top(\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R}^\top)^{-1}\left[\boldsymbol{R}\hat{\boldsymbol{\beta}}^{\mathrm{SUS}} - \boldsymbol{r}\right]\right]. \tag{3.2}$$

Note, Proposition 3.2 gives the risk of the restricted Liu machine learner defined in Proposition 3.1. After some algebra, one may see that the risk of the weighted WLSE, i.e., $\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{\text{SUS}}(\boldsymbol{K})$ is smaller. Hence, in the sequel, we use (3.2) for the forthcoming developments. Note, that the expression inside the curly bracket of (3.2) is the restricted WLSE. It can be obtained in the same fashion as in the proof of Proposition 3.1.

3.1 Development of a class of shrinkage learners

Besides the computation of the LWLSE restricted learner, it is desirable to check whether the set of restrictions $\boldsymbol{R}_i\boldsymbol{\beta}_i = \boldsymbol{r}_i$, $i = 1, \dots, M$ is true. The usage of the LWLSE is highly dependent on the correctness of the null hypothesis. Formulating Wlad's test statistics is straightforward using the asymptotic distribution of the WLSE. Here, we develop a class of combinations of the LWLSE and its restricted version to calibrate the effect of the set of restrictions.

For our purpose, consider the general class of shrinkage learners given by

$$\hat{\boldsymbol{\beta}}^{\text{Shrinkage}}(\boldsymbol{K}) = \hat{\boldsymbol{\beta}}^{\text{SUS}}(\boldsymbol{K}) - \left(\hat{\boldsymbol{\beta}}^{\text{SUS}}(\boldsymbol{K}) - \hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{\text{SUS}}(\boldsymbol{K})\right) g(T_n), \tag{3.3}$$

where T_n is a proposed test statistic for testing $\boldsymbol{R\beta} = \boldsymbol{r}$. Table 1 gives the formulation of three shrinkage learners. Apparently, the linear shrinkage machine learner is obtained by taking $g(T_n) = w$ for some $w \in [0, 1]$.

Table 1: Expressions for the corresponding $T_n(\cdot)$ functions in the proposed shrinkage learners. Here, t is the critical value for the test statistic, and $w \in [0, 1]$ is a constant.

Shrinkage learner	$g(\cdot)$ function	Designation
$\hat{\boldsymbol{\beta}}^{\text{PT}-\text{SUS}}(\boldsymbol{K})$	$I\left(T_n \le t\right)$	Preliminary-test
$\hat{\boldsymbol{\beta}}^{\text{S}-\text{SUS}}(\boldsymbol{K})$	$(nM-2)T_n^{-1}$	Stein-type
$\hat{\boldsymbol{\beta}}^{\text{PR}-\text{SUS}}(\boldsymbol{K})$	$(1-(nM-2)T_n^{-1})I(T_n \le t)$	Positive-rule Stein-type

Using the results of Roozbeh et al. (2012), the WLSE is asymptotically normally distributed with mean $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{C}^{-1}$. In the case of unknown $\boldsymbol{V}$, we use $\boldsymbol{S}$, and we get the same normality distribution by Zellner (1962). Therefore, for testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{R\beta} = \boldsymbol{r}$, we use the following test statistic:

$$T_n = \frac{1}{nM}(\boldsymbol{R}\hat{\boldsymbol{\beta}}^{\text{SUS}} - \boldsymbol{r})^\top(\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R})^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}^{\text{SUS}} - \boldsymbol{r}). \tag{3.4}$$

Then, under the null hypothesis, T_n has an asymptotic χ^2 distribution with nM degrees of freedom (d.f.). In case of the unknown $\boldsymbol{V}$, by replacing $\boldsymbol{V}$ with $\boldsymbol{S}$, T_n has an asymptotic F-distribution with $(nM, p-m)$ d.f. Thus, in Table 1, $t = \chi^2_{\alpha,p}$ for $\alpha \in (0, 1)$. Under an alternative hypothesis, we have a non-central distribution with a non-centrality parameter:

$$\Delta = (\boldsymbol{R\beta} - \boldsymbol{r})^\top(\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R})^{-1}(\boldsymbol{R\beta} - \boldsymbol{r}).$$

3.2 Theoretical characteristics

The proposed restricted Liu machine learner $\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{\mathrm{SUS}}(\boldsymbol{K})$ possesses superior performance over its counterpart, i.e., $\hat{\boldsymbol{\beta}}^{\mathrm{SUS}}$. We refer to Roozbeh et al. (2012) for a similar approach and do not provide further details to avoid repetition. Here, we mainly focus on shrinkage versions of the LWLSE restricted learner. Because of our proposal's restricted nature, we study the analytical performance for the class of local alternatives $\{K_{(n,M)}\}$ given by

$$K_{(n,M)} : \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r} + (nM)^{-\frac{1}{2}}\boldsymbol{\xi}. \tag{3.5}$$

Note that here, the effective sample dimension is nM. Refer to Saleh et al. (2019) and Saleh et al. (2022) for details. Hence, asymptotic distributional properties shall be studied first. To this end, we compute the asymptotic bias and risk in terms of distribution.

Now, assume for any estimator $\hat{\boldsymbol{\beta}}^*$ of $\boldsymbol{\beta}$, the asymptotic cumulative distribution exists and is given by $F(\boldsymbol{x}) = \lim\limits_{n\to\infty} \Pr(\sqrt{nM}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) \leq \boldsymbol{x})$. Then, the asymptotic distributional bias and risk, respectively, are given by

$$\mathrm{Bias}(\hat{\boldsymbol{\beta}}) = \lim_{n\to\infty} \mathbb{E}\left(n^{\frac{1}{2}}M^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right),$$
$$\mathrm{Risk}(\hat{\boldsymbol{\beta}}) = \mathrm{tr}\left(\int_{\mathbb{R}^p} \boldsymbol{x}\boldsymbol{x}^\top \mathrm{d}F(\boldsymbol{x})\right) = \mathrm{tr}(\boldsymbol{V}),$$

where $\boldsymbol{V}$ is the dispersion matrix for the distribution $F(\boldsymbol{x})$.

Proposition 3.2. The asymptotic distributional bias and associated risk of the learner in Proposition 3.1 are, respectively, given by

$$\mathrm{Bias} = \boldsymbol{M_K}(\boldsymbol{K} - \boldsymbol{I}_p)\boldsymbol{\beta}, \quad \text{and} \quad \mathrm{Risk} = \mathrm{tr}\left(\boldsymbol{M_K}(\boldsymbol{K}+\boldsymbol{C})\boldsymbol{C}^{-1}(\boldsymbol{K}+\boldsymbol{C})\boldsymbol{M_K}\right),$$

where $\boldsymbol{M_K} = (\boldsymbol{C}+\boldsymbol{I}_p)^{-1} - (\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top(\boldsymbol{R}(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top)^{-1}\boldsymbol{R}(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}$.

For the sketch of proof, refer to the Appendix. Note that we do not use Proposition 3.2 in the sequel since the focus is on $\hat{\boldsymbol{\beta}}^{\mathrm{SUS}}(\boldsymbol{K})$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{\mathrm{SUS}}(\boldsymbol{K})$.

In an asymptotic sense, for the shrinkage methodologies, it is easy to see

$$\sqrt{nM}(\hat{\boldsymbol{\beta}}^{\mathrm{SUS}}(\boldsymbol{K}) - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N_p\left((\boldsymbol{T_K} - \boldsymbol{I}_p)\boldsymbol{\beta}, \boldsymbol{C_K}\right), \tag{3.6}$$
$$\sqrt{nM}(\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{\mathrm{SUS}}(\boldsymbol{K}) - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N_p\left(\boldsymbol{M_K}(\boldsymbol{K} - \boldsymbol{I}_p)\boldsymbol{\beta}, \boldsymbol{D_K}\right), \tag{3.7}$$

where $\boldsymbol{C_K} = \boldsymbol{T_K}\boldsymbol{C}^{-1}\boldsymbol{T_K}$ and $\boldsymbol{D_K} = \boldsymbol{M_K}(\boldsymbol{K}+\boldsymbol{C})\boldsymbol{C}^{-1}(\boldsymbol{K}+\boldsymbol{C})\boldsymbol{M_K}$.

The bias expressions of the proposed estimators are directly adapted from Saleh (2006). Hence, we omit the detailed derivations and instead provide the bias formulas in the following theorem.

Proposition 3.3. The asymptotic distributional biases of all Liu machine learners are given by

$$\mathrm{Bias}(\hat{\boldsymbol{\beta}}^{\mathrm{SUS}}(\boldsymbol{K})) = -(\boldsymbol{I}_p - \boldsymbol{K})(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{\beta};$$
$$\mathrm{Bias}(\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{\mathrm{SUS}}(\boldsymbol{K})) = -\left[(\boldsymbol{I}_p - \boldsymbol{K})(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{\beta} + \boldsymbol{T_K}\boldsymbol{\eta}\right];$$
$$\mathrm{Bias}(\hat{\boldsymbol{\beta}}^{\mathrm{PT-SUS}}(\boldsymbol{K})) = -\left[(\boldsymbol{I}_p - \boldsymbol{K})(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{\beta} + \boldsymbol{T_K}\boldsymbol{\eta}H_{p+2}(\chi^2_{\alpha,p};\Delta)\right];$$

$$\text{Bias}(\hat{\boldsymbol{\beta}}^{\text{S-SUS}}(\boldsymbol{K})) = -\left[(\boldsymbol{I}_p - \boldsymbol{K})(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{\beta} + (p-2)\boldsymbol{T}_{\boldsymbol{K}}\boldsymbol{\eta}\mathbb{E}(\chi_{p+2}^{-2}(\Delta))\right];$$
$$\begin{aligned}\text{Bias}(\hat{\boldsymbol{\beta}}^{\text{PR-SUS}}(\boldsymbol{K})) = -\Big[&(\boldsymbol{I}_p - \boldsymbol{K})(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{\beta} + \boldsymbol{T}_{\boldsymbol{K}}\boldsymbol{\eta}\Big\{(p-2)\mathbb{E}^{(2)}[\chi_{p+2}^{-2}(\Delta)I(\chi_{p+2}^{2}(\Delta) < p-2)]\\ &- (p-2)\mathbb{E}^{(2)}(\chi_{p+2}^{-2}(\Delta)I(\chi_{p+2}^{2}(\Delta) < p-2)) - H_{p+2}(\chi_{\alpha,p}^{2};\Delta)\Big\}\Big],\end{aligned}$$

where $\boldsymbol{\eta} = \boldsymbol{\delta}\boldsymbol{\xi}$, $\boldsymbol{\delta} = \boldsymbol{C}^{-1}\boldsymbol{R}^{\top}(\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R}^{\top})^{-1}$, and $H_q(x;\Delta)$ is the cumulative distribution function of a non-central χ-square distribution with q d.f. and non-centrality parameter Δ. Furthermore,

$$\mathbb{E}(\chi_q^{-2i}(\Delta)) = \int_{x=0}^{x=\infty} x^{-2i} dH_q(x;\Delta), \quad i = 1, 2.$$

For $\alpha = 0$, the bias of $\hat{\boldsymbol{\beta}}^{\text{PT-SUS}}(\boldsymbol{K})$ coincides with that of the restricted Liu learner $\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{\text{SUS}}(\boldsymbol{K})$, while for $\alpha = 1$, it coincides with that of $\hat{\boldsymbol{\beta}}^{\text{SUS}}(\boldsymbol{K})$, the unrestricted Liu learner of $\boldsymbol{\beta}$. Also, as the departure parameter $\Delta \to \infty$, we have:

$$\text{Bias}(\hat{\boldsymbol{\beta}}^{\text{PT-SUS}}(\boldsymbol{K})) = \text{Bias}(\hat{\boldsymbol{\beta}}^{\text{S-SUS}}(\boldsymbol{K})) = \text{Bias}(\hat{\boldsymbol{\beta}}^{\text{PR-SUS}}(\boldsymbol{K})) = \text{Bias}(\hat{\boldsymbol{\beta}}^{\text{SUS}}(\boldsymbol{K})),$$

while $\text{Bias}(\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{\text{SUS}}(\boldsymbol{K}))$ becomes unbounded. Under $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{r}$, all the estimators are inherently biased, and the magnitude of this bias is identical across all the proposed estimators.

Proposition 3.4. The asymptotic distributional risks of all Liu machine learners are given by

$$\begin{aligned}
\text{Risk}(\hat{\boldsymbol{\beta}}^{\text{SUS}}(\boldsymbol{K})) &= \text{tr}(\boldsymbol{T}_{\boldsymbol{K}}^{\top}\boldsymbol{C}^{-1}\boldsymbol{T}_{\boldsymbol{K}}) + \boldsymbol{\beta}^{\top}(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}(\boldsymbol{I}_p-\boldsymbol{K})^2(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{\beta};\\
\text{Risk}(\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{\text{SUS}}(\boldsymbol{K})) &= \text{tr}(\boldsymbol{T}_{\boldsymbol{K}}^{\top}\boldsymbol{C}^{-1}\boldsymbol{T}_{\boldsymbol{K}}) - \text{tr}(\boldsymbol{T}_{\boldsymbol{K}}^{\top}\boldsymbol{A}\boldsymbol{T}_{\boldsymbol{K}}) + \boldsymbol{\eta}^{\top}\boldsymbol{T}_{\boldsymbol{K}}^{\top}\boldsymbol{T}_{\boldsymbol{K}}\boldsymbol{\eta}\\
&\quad + 2\boldsymbol{\eta}^{\top}\boldsymbol{T}_{\boldsymbol{K}}^{\top}(\boldsymbol{I}_p-\boldsymbol{K})(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{\beta} + \boldsymbol{\beta}^{\top}(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}(\boldsymbol{I}_p-\boldsymbol{K})\boldsymbol{\beta};\\
\text{Risk}(\hat{\boldsymbol{\beta}}^{\text{PT-SUS}}(\boldsymbol{K})) &= \text{tr}(\boldsymbol{T}_{\boldsymbol{K}}^{\top}\boldsymbol{C}^{-1}\boldsymbol{T}_{\boldsymbol{K}}) - \text{tr}(\boldsymbol{T}_{\boldsymbol{K}}^{\top}\boldsymbol{A}\boldsymbol{T}_{\boldsymbol{K}})H_{p+2}(\chi_{\alpha,p}^2;\Delta)\\
&\quad + \boldsymbol{\eta}^{\top}\boldsymbol{T}_{\boldsymbol{K}}^{\top}\boldsymbol{T}_{\boldsymbol{K}}\boldsymbol{\eta}Z(\alpha,\Delta) + 2\boldsymbol{\eta}^{\top}\boldsymbol{T}_{\boldsymbol{K}}^{\top}(\boldsymbol{I}_p-\boldsymbol{K})(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{\beta}H_{p+2}(\chi_{\alpha,p}^2;\Delta)\\
&\quad + \boldsymbol{\beta}^{\top}(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}(\boldsymbol{I}_p-\boldsymbol{K})^2(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{\beta};\\
\text{Risk}(\hat{\boldsymbol{\beta}}^{\text{S-SUS}}(\boldsymbol{K})) &= \text{tr}(\boldsymbol{T}_{\boldsymbol{K}}^{\top}\boldsymbol{C}^{-1}\boldsymbol{T}_{\boldsymbol{K}}) - (p-2)\text{tr}(\boldsymbol{T}_{\boldsymbol{K}}^{\top}\boldsymbol{A}\boldsymbol{T}_{\boldsymbol{K}})X(\Delta)\\
&\quad + (p-2)\boldsymbol{\eta}^{\top}\boldsymbol{T}_{\boldsymbol{K}}^{\top}\boldsymbol{T}_{\boldsymbol{K}}\boldsymbol{\eta}Y(\Delta) + 2(p-2)\boldsymbol{\eta}^{\top}\boldsymbol{T}_{\boldsymbol{K}}^{\top}(\boldsymbol{I}_p-\boldsymbol{K})(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{\beta}\mathbb{E}(\chi_{p+2}^{-2}(\Delta))\\
&\quad + \boldsymbol{\beta}^{\top}(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}(\boldsymbol{I}_p-\boldsymbol{K})^2(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{\beta};\\
\text{Risk}(\hat{\boldsymbol{\beta}}^{\text{PR-SUS}}(\boldsymbol{K})) &= \text{Risk}(\hat{\boldsymbol{\beta}}^{\text{S-SUS}}(\boldsymbol{K}))\\
&\quad - \left\{\text{tr}(\boldsymbol{T}_{\boldsymbol{K}}^{\top}\boldsymbol{A}\boldsymbol{T}_{\boldsymbol{K}})\mathbb{E}\left[(1-(p-2)\chi_{p+2}^{-2}(\Delta))^2 I(\chi_{p+2}^2(\Delta) < p-2)\right]\right.\\
&\quad \left. + \boldsymbol{\eta}^{\top}\boldsymbol{T}_{\boldsymbol{K}}^{\top}\boldsymbol{T}_{\boldsymbol{K}}\boldsymbol{\eta}\mathbb{E}\left[(1-(p-2)\chi_{p+2}^{-2}(\Delta))^2 I(\chi_{p+4}^2(\Delta) < p-2)\right]\right\}\\
&\quad - 2\boldsymbol{\eta}^{\top}\boldsymbol{T}_{\boldsymbol{K}}^{\top}\boldsymbol{T}_{\boldsymbol{K}}\boldsymbol{\eta}\mathbb{E}\left[((p-2)\chi_{p+2}^{-2}(\Delta)-1)I(\chi_{p+2}^2(\Delta) < p-2)\right]\\
&\quad - 2\boldsymbol{\eta}^{\top}\boldsymbol{T}_{\boldsymbol{K}}^{\top}(\boldsymbol{I}_p-\boldsymbol{K})(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{\beta}\mathbb{E}\left[((p-2)\chi_{p+2}^{-2}(\Delta)-1)I(\chi_{p+2}^2(\Delta) < p-2)\right],
\end{aligned}$$

where

$$\begin{aligned}
X(\Delta) &= 2\mathbb{E}(\chi_{p+2}^{-2}(\Delta)) - (p-2)\mathbb{E}(\chi_{p+2}^{-4}(\Delta));\\
Y(\Delta) &= 2\mathbb{E}(\chi_{p+2}^{-2}(\Delta)) - 2\mathbb{E}(\chi_{p+4}^{-2}(\Delta)) + (p-2)\mathbb{E}(\chi_{p+4}^{-4}(\Delta));\\
Z(\alpha,\Delta) &= 2H_{p+2}(\chi_{\alpha,p}^2;\Delta) - H_{p+4}(\chi_{\alpha,p}^2;\Delta),
\end{aligned}$$

and $\boldsymbol{A} = \boldsymbol{C}^{-1}\boldsymbol{R}^{\top}(\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{r}^{\top})^{-1}\boldsymbol{R}\boldsymbol{C}^{-1}$.

4 Conclusion

This paper delved into only the theoretical construction methodology of widely recognized shrinkage estimators, placing special emphasis on Liu estimators, which exhibit exceptional performance in scenarios characterized by multicollinearity. The phenomenon of multicollinearity arises when predictor variables in a regression model exhibit high correlations, leading to instability in the estimation of coefficients. This instability often manifests as inflated standard errors, making it challenging to determine the individual impact of each predictor variable. Consequently, the reliability and interpretability of statistical analyses are compromised. In addressing these challenges, Liu estimators provide a compelling solution by introducing a penalization mechanism that effectively controls bias while improving the efficiency of parameter estimates. The shrinkage Liu-type learner is particularly advantageous as it not only mitigates the adverse effects of multicollinearity but also enhances the robustness of estimates, allowing for more accurate inference.

The methodology for constructing these shrinkage learners builds on a solid theoretical foundation, integrating principles from both classical regression techniques and modern statistical best practices. By carefully formulating the estimator in relation to the underlying data structure, we can tailor applications to specific regression contexts, thereby maximizing performance. Moreover, the proposed Liu estimators can be readily implemented across various statistical software platforms, facilitating their application in diverse fields such as economics, biostatistics, and social sciences. The resulting implementation is designed to be user-friendly, providing practitioners with the necessary tools to apply these advanced methodologies without extensive background knowledge in statistical theory.

Through this exploration, we aimed to contribute to the broader discourse on shrinkage methods in regression analysis, equipping researchers and practitioners with a deeper understanding of Liu estimators and their potential applications.

Acknowledgements

We thank Professor B.M. Golam Kibria for the invitation to contribute to a special issue honoring late Professor A. K. Md. Ehsanes Saleh. Mohammad Arashi's work is based on the research supported in part by the Iran National Science Foundation (INSF) grant No. 4015320.

References

Abonazel, M. R. (2023), "New modified two-parameter Liu estimator for the Conway–Maxwell Poisson regression model," *Journal of Statistical Computation and Simulation*, 93, 1976–1996.

Akdeniz, F. and Kaciranlar, S. (1995), "On the almost unbiased generalized Liu estimator and unbiased estimation of the bias and MSE," *Comm. Statist. Theo. Meth.*, 24, 1789–1797.

Akram, M. N., Amin, M., Kibria, B. M. G., Arashi, M., Lukman, A. F., and Afzal, N. (2024), "A new improved Liu estimator for the QSAR model with inverse Gaussian response," *Communications in Statistics - Simulation and Computation*, 53, 1873–1888.

Al-Momani, M. (2023), "Liu-type pretest and shrinkage estimation for the conditional autoregressive model," *PLOS ONE*, 18, e0283339.

Algamal, Z. Y. and Abonazel, M. R. (2022), "Developing a Liu-type estimator in beta regression model," *Concurrency Computation: Practice and Experience*, 34, e6685.

Altukhaes, W. B., Roozbeh, M., and Mohamed, N. A. (2024a), "Feasible robust Liu estimator to combat outliers and multicollinearity effects in restricted semiparametric regression model," *AIMS Mathematics*, 9, 31581–31606.

— (2024b), "Robust Liu estimator used to combat some challenges in partially linear regression model by improving LTS algorithm using semidefinite programming," *Mathematics*, 12, 2787.

Amin, M., Akram, M. N., and Kibria, B. M. G. (2021), "A new adjusted Liu estimator for the Poisson regression model," *Concurrency Computation: Practice and Experience*, 33, e6340.

Arashi, M., Lukman, A. F., and Algamal, Z. Y. (2022), "Liu regression after random forest for prediction and modeling in high-dimension," *Journal of Chemometrics*, 36, e3393.

Arashi, M., Nadarajah, S., and Akdeniz, F. (2017), "The distribution of the Liu type estimator of the biasing parameter in elliptically contoured models," *Communication in Statistics – Theory and Methods*, 46, 3829–3837.

Arashi, M., Norouzirad, M., Ahmed, S. E., and Yuzbasi, B. (2018), "Rank-based Liu regression," *Computational Statistics*, 33, 1525–1561.

Asar, Y., Erisoglu, M., and Arashi, M. (2017), "Developing a restricted two-parameter Liu-type estimator: A comparison of restricted estimators in the binary logistic regression model," *Communication in Statistics – Theory and Methods*, 46, 6864–6873.

Baltagi, B. (1980), "On seemingly unrelated regressions with error components," *Econometrica*, 48, 1547–1552.

Chib, S. and Greenberg, E. (1995), "Hierarchical analysis of SUR model with extensions to correlated serial errors and time varying parameter models," *Journal of Econometrics*, 68, 339–360.

Fiebig, D. G. (2001), "Seemingly Unrelated Regression," in *A Companion to Theoretical Econometrics*, ed. Baltagi, B., Blackwell Publishers, pp. 101–121.

Genç, M. and Lukman, A. (2025), "Weighted LAD-Liu-LASSO for robust estimation and sparsity," *Computational Statistics*.

Ghanem, E., Hatefi, A., and Usefi, H. (2024), "Liu-type shrinkage estimators for mixture of logistic regressions: an osteoporosis study," *Communications in Statistics - Simulation and Computation*, 1–15.

Hawa, N. S., Mustafa, M. Y., Kibria, B. M. G., and Algamal, Z. Y. (2025), "Bootstrap Liu-type estimator for Conway-Maxwell-Poisson regression model," *Communications in Statistics - Simulation and Computation*, 1–12.

Karlsson, P., Månsson, K., and Kibria, B. M. G. (2020), "A Liu estimator for the beta regression model and its application to chemical data," *Journal of Chemometrics*, 34, e3300.

Kibria, B. M. G. (2003), "Performance of Some New Ridge Regression Estimators," *Communications in Statistics - Simulation and Computation*, 32, 419–435.

— (2022), "More than hundred (100) estimators for estimating the shrinkage parameter in a linear and generalized linear ridge regression models," *Journal of Econometrics and Statistics*, 2, 233–252.

Liu, K. (1993), "A new class of biased estimate in linear regression," *Comm. Statist. Theo. Meth.*, 22, 393–402.

Lukman, A. F., Kibria, B. M. G., Ayinde, K., and Jegede, S. L. (2020), "Modified One-Parameter Liu Estimator for the Linear Regression Model," *Modelling and Simulation in Engineering*, 9574304.

Moon, H. R. (1999), "A note on fully-modified estimation of seemingly unrelated regressions models with integrated regressors," *Economics Letters*, 65, 25–31.

Moon, H. R. and Perron, P. (2004), "Efficient estimation of SUR cointegration regression model and testing for purchasing power parity," *Economic Review*, 23, 293–323.

Qasim, M., Kibria, B. M. G., Månsson, K., and Sjölander, P. (2019), "A new Poisson Liu Regression Estimator: method and application," *Journal of Applied Statistics*, 47, 2258–2271.

Roozbeh, M. and Arashi, M. (2014), "Feasible ridge estimator in seemingly unrelated semiparametric models," *Communications in Statistics - Simulation and Computation*, 116, 35–44.

Roozbeh, M., Arashi, M., and Niroumand, H. A. (2012), "Seemingly unrelated ridge regression in semiparametric models," *Communications in Statistics - Theory and Methods*, 41, 1364–1386.

Saleh, A. K. M. E., Arashi, M., and Kibria, B. M. G. (2019), *Theory of Ridge Regression Estimation with Applications*, USA: Wiley.

Saleh, A. K. M. E., Arashi, M., Saleh, R. A., and Norouzirad, M. (2022), *Rank-Based Methods for Shrinkage and Selection: With Application to Machine Learning*, USA: Wiley.

Speckman, P. (1988), "Kernel smoothing in partial linear models," *J. Royal Statist. Soc.: Ser. B*, 50, 413–437.

Srivastava, V. K. and Giles, D. E. A. (1987), *Seemingly Unrelated Regression Equation Models*, Marcel Dekker.

Srivastava, V. K. and Maekawa, K. (1995), "Efficiency properties of feasible generalized least squares estimators in SURE models under non-normal disturbances," *Journal of Econometrics*, 66, 99–121.

Tanış, C. and Asar, Y. (2024), "Liu-type estimator in Conway–Maxwell–Poisson regression model: theory, simulation and application," *Statistics*, 58, 65–86.

Wu, J., Asar, Y., and Arashi, M. (2018), "On the restricted almost unbiased Liu estimator in the Logistic regression model," *Communication in Statistics – Theory and Methods*, 47, 4389–4401.

Zellner, A. (1962), "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," *Journal of the American Statistical Association*, 57, 348–368.

Received: March 7, 2025

Accepted: June 16, 2025

A Appendix

In this section, we give the proof of main results along with some technical lemmas.
Proof of Proposition 3.1: To formulate the Liu learner with restrictions, we consider the following constrained optimization problem

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}(\widetilde{\boldsymbol{Y}}-\widetilde{\boldsymbol{X}}\boldsymbol{\beta})^\top\boldsymbol{V}^{-1}(\widetilde{\boldsymbol{Y}}-\widetilde{\boldsymbol{X}}\boldsymbol{\beta}),\ \text{s.t. } (\boldsymbol{K}\hat{\boldsymbol{\beta}}^{\text{SUS}}-\boldsymbol{\beta})^\top(\boldsymbol{K}\hat{\boldsymbol{\beta}}^{\text{SUS}}-\boldsymbol{\beta})\leq\rho^2 \text{ and } \boldsymbol{R}\boldsymbol{\beta}=\boldsymbol{r}.$$

Using the Lagrangian multipliers, we solve

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}(\widetilde{\boldsymbol{Y}}-\widetilde{\boldsymbol{X}}\boldsymbol{\beta})^\top\boldsymbol{V}^{-1}(\widetilde{\boldsymbol{Y}}-\widetilde{\boldsymbol{X}}\boldsymbol{\beta})+(\boldsymbol{K}\hat{\boldsymbol{\beta}}^{\text{SUS}}-\boldsymbol{\beta})^\top(\boldsymbol{K}\hat{\boldsymbol{\beta}}^{\text{SUS}}-\boldsymbol{\beta})+2\boldsymbol{\lambda}_2^\top(\boldsymbol{R}\boldsymbol{\beta}-\boldsymbol{r}),$$

where $\boldsymbol{\lambda}_2$ is the set of Lagrangian coefficients. Getting derivative concerning $\boldsymbol{\beta}$ and equating to zero, yield

$$(\widetilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}}+\boldsymbol{I}_p)\boldsymbol{\beta}=\widetilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-1}\widetilde{\boldsymbol{Y}}+\boldsymbol{K}\hat{\boldsymbol{\beta}}^{\text{SUS}}-\boldsymbol{R}^\top\boldsymbol{\lambda}_2.$$

Therefore, we get

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{SUS}(\boldsymbol{K})&=(\widetilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-1}\widetilde{\boldsymbol{X}}+\boldsymbol{I}_p)^{-1}\left[\widetilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-1}\widetilde{\boldsymbol{Y}}+\boldsymbol{K}\hat{\boldsymbol{\beta}}^{\text{SUS}}-\boldsymbol{R}^\top\boldsymbol{\lambda}_2\right]\\&=(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}(\boldsymbol{C}+\boldsymbol{K})\hat{\boldsymbol{\beta}}^{\text{SUS}}-(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top\boldsymbol{\lambda}_2\\&=\boldsymbol{T}_{\boldsymbol{K}}\hat{\boldsymbol{\beta}}^{\text{SUS}}-(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top\boldsymbol{\lambda}_2\end{aligned}\tag{A.1}$$

Using the fact that $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$, premultiplying (A.1) with $\boldsymbol{R}$ and equating with $\boldsymbol{r}$ results in

$$\boldsymbol{R}\left\{\boldsymbol{T_K}\hat{\boldsymbol{\beta}}^{\mathrm{SUS}} - (\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top\boldsymbol{\lambda}_2\right\} = \boldsymbol{r}.$$

Solving the above equality for $\boldsymbol{\lambda}_2$ gives

$$\boldsymbol{\lambda}_2 = (\boldsymbol{R}(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top)^{-1}\left[\boldsymbol{R}\boldsymbol{T_K}\hat{\boldsymbol{\beta}}^{\mathrm{SUS}} - \boldsymbol{r}\right]$$

Substituting $\boldsymbol{\lambda}_2$ in (A.1) yields

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{SUS}(\boldsymbol{K}) &= \boldsymbol{T_K}\hat{\boldsymbol{\beta}}^{\mathrm{SUS}} - (\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top(\boldsymbol{R}(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top)^{-1}\left[\boldsymbol{R}\boldsymbol{T_K}\hat{\boldsymbol{\beta}}^{\mathrm{SUS}} - \boldsymbol{r}\right]\\ &= \hat{\boldsymbol{\beta}}^{\mathrm{SUS}}(\boldsymbol{K}) - (\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top(\boldsymbol{R}(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top)^{-1}\left[\boldsymbol{R}\hat{\boldsymbol{\beta}}^{\mathrm{SUS}}(\boldsymbol{K}) - \boldsymbol{r}\right].\end{aligned}$$

The proof is complete.
Proof of Proposition 3.2: Define

$$\begin{aligned}\boldsymbol{\beta}_0 &= \boldsymbol{R}^\top(\boldsymbol{R}\boldsymbol{R}^\top)^{-1}\boldsymbol{r}\\ \boldsymbol{M_K} &= (\boldsymbol{C}+\boldsymbol{I}_p)^{-1} - (\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top(\boldsymbol{R}(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top)^{-1}\boldsymbol{R}(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}.\end{aligned}$$

Then, by Proposition 3.1, it is easy to see that

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{SUS}(\boldsymbol{K}) &= \hat{\boldsymbol{\beta}}^{\mathrm{SUS}}(\boldsymbol{K}) - (\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top(\boldsymbol{R}(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top)^{-1}\left[\boldsymbol{R}\hat{\boldsymbol{\beta}}^{\mathrm{SUS}}(\boldsymbol{K}) - \boldsymbol{r}\right]\\ &= \boldsymbol{M_K}(\boldsymbol{C}+\boldsymbol{I}_p)\hat{\boldsymbol{\beta}}^{\mathrm{SUS}}(\boldsymbol{K}) - \boldsymbol{M_K}(\boldsymbol{C}+\boldsymbol{I}_p)\boldsymbol{\beta}_0 + \boldsymbol{\beta}_0\\ &= \boldsymbol{M_K}(\boldsymbol{K}\boldsymbol{C}^{-1}+\boldsymbol{I}_p)\widetilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-1}\widetilde{\boldsymbol{Y}} - \boldsymbol{M}_k(\boldsymbol{C}+\boldsymbol{I}_p)\boldsymbol{\beta}_0 + \boldsymbol{\beta}_0. \qquad \text{(A.2)}\end{aligned}$$

Therefore, using the fact that $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r} = \boldsymbol{R}\boldsymbol{\beta}_0$, we get

$$\begin{aligned}\mathrm{Bias}(\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{SUS}(\boldsymbol{K})) &= \boldsymbol{M_K}(\boldsymbol{K}\boldsymbol{C}^{-1}+\boldsymbol{I}_p)\boldsymbol{C}\boldsymbol{\beta} - \boldsymbol{M_K}(\boldsymbol{C}+\boldsymbol{I}_p)\boldsymbol{\beta}_0 + \boldsymbol{\beta}_0 - \boldsymbol{\beta}\\ &= \boldsymbol{M_K}(\boldsymbol{K}-\boldsymbol{I}_p+\boldsymbol{I}_p+\boldsymbol{C})\boldsymbol{\beta} - \boldsymbol{M_K}(\boldsymbol{C}+\boldsymbol{I}_p)\boldsymbol{\beta}_0 + \boldsymbol{\beta}_0 - \boldsymbol{\beta}\\ &= \boldsymbol{M_K}(\boldsymbol{C}+\boldsymbol{I}_p)(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + \boldsymbol{M_K}(\boldsymbol{K}-\boldsymbol{I}_p)\boldsymbol{\beta} + \boldsymbol{\beta}_0 - \boldsymbol{\beta}\\ &= \left\{\boldsymbol{I}_p - (\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top(\boldsymbol{R}(\boldsymbol{C}+\boldsymbol{I}_p)^{-1}\boldsymbol{R}^\top)^{-1}\boldsymbol{R}\right\}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\\ &\quad + \boldsymbol{M_K}(\boldsymbol{K}-\boldsymbol{I}_p)\boldsymbol{\beta} + \boldsymbol{\beta}_0 - \boldsymbol{\beta}\\ &= \boldsymbol{M_K}(\boldsymbol{K}-\boldsymbol{I}_p)\boldsymbol{\beta}.\end{aligned}$$

Also, by definition

$$\begin{aligned}\mathrm{Risk}(\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}^{SUS}(\boldsymbol{K})) &= \mathrm{trCov}(\hat{\boldsymbol{\beta}}_{\mathrm{r}}^{\mathrm{SUS}}(\mathrm{K}))\\ &= \mathrm{tr}\left(\boldsymbol{M_K}(\boldsymbol{K}+\boldsymbol{C})\boldsymbol{C}^{-1}(\boldsymbol{K}+\boldsymbol{C})\boldsymbol{M_K}\right).\end{aligned}$$