Comput Stat (2014) 29:1515–1541
DOI 10.1007/s00180-014-0504-3

ORIGINAL PAPER

A VIF-based optimization model to alleviate collinearity problems in multiple linear regression

Yow-Jen Jou · Chien-Chia Liäm Huang · Hsun-Jung Cho

Received: 2 October 2012 / Accepted: 9 May 2014 / Published online: 5 June 2014

Abstract In this paper, we address data collinearity problems in multiple linear regression from an optimization perspective. We propose a novel linearly constrained quadratic programming model, based on the concept of the variance inflation factor (*VIF*). We employ the perturbation method that involves imposing a general symmetric non-diagonal perturbation matrix on the correlation matrix. The proposed *VIF*-based model reduces the largest *VIF* by minimizing the resulting biases. The *VIF*-based model can mitigate the harm from data collinearity through the reduction in both the condition number and *VIF*s, meanwhile improving the statistical significance. The resulting estimator has bounded biases under an iterative framework and hence is termed the *least accumulative bias estimator*. Certain potential statistical properties can be further considered as the side constraints for the proposed model. Various numerical examples validate the proposed approach.

Keywords Multicollearity · Variance inflation factor · Convex optimization

Y.-J. Jou
Department of Information Management and Finance, National Chiao Tung University,
Room 422, Management Building , Hsinchu 30010, Taiwan
e-mail: yjjou.dif@gmail.com

C.-C. L. Huang · H.-J. Cho
Department of Transportation and Logistics Management, National Chiao Tung University,
814 General Building, Hsinchu 30010, Taiwan
e-mail: hjcho001@gmail.com

C.-C. L. Huang (✉)
Department of Industrial and Systems Engineering, North Carolina State University,
Room 443, Daniels Hall, NCSU Campus, Raleigh, NC 27606, USA
e-mail: chuang10@ncsu.edu

1 Introduction

Motivation Collinearity problem describes the situation that the non-orthogonality exists among explanatory variables in regression models. The breakdown in the orthogonality among explanatory variables causes imprecisions in the use of normal equation in the ordinary least squares (OLSs) estimations. The imprecision oftentimes leads to high variances, and thus low statistical significance, and even the incorrect signs of the OLS estimators (Shen and Wohlgenant 2010). This renders the OLS estimators, even though they still exist, inappropriate in statistical senses. Moreover, the power in prediction of the established regression model is far weakened (Belsley 1984). Hence, collinearity deserves more attention because its presence has frustrated the researchers endeavoring to establish important relationship among interesting variables using regression models (Shacham and Brauner 1997; Næs and Mevik 2001).

Belsley (1980) ascribed collinearity problem to a *data problem*, instead of a statistical one. This is because the specification of the models has implied the independence among the explanatory variables. Inevitably, yet, we work with non-experimental data, pooling from an unknown vast population. Uncertainties and unpredictabilities in the behaviour of the collected samples oftentimes lead to unsatisfactory results. Much effort has been sowed in designing sophisticated experiments as well as new techniques to, at least, mitigate the harm from data collinearity, for example the ridge regression (Hoerl and Kennard 1970), the LASSO (Tibshirani 1996) and the bridge regression (Frank and Friedman 1993).

Spanos and McGuirk (2002) discussed the near-multicollinearity problems by classifying the problems into (i) a structual issue (*systematic volatility*) and (ii) a numerical issue (*erratic volatility*). Systematic volatility refers to high correlations among the explanatory variables, concerning the structure of the correlation matrix that potentially invokes the presence of the data collinearity. The numerical issues concern the ill-conditionedness of the data matrix.

Ridge regression has won its reputation in addressing data collinearity problems successfully by imposing perturbations on the diagonals of a correlation matrix. Ridge regression takes effect by sacrificing the intrinsic structure of the correlation matrix, viz., the correlation between a variable and itself is 1 by definition. The perturbation, no matter how small, on the diagonals of the correlation matrix violates this definition essentially.

We aim at the proposal of a novel approach that incorporates both the systematic and erratic volatilities, as aformentioned. From the systematic perspective, the proposed approach mitigates the effects from data collinearity by preserving the intrinsic structure of the correlation matrix, in which aspect the ridge regression does not. From the erratic perspective, the proposed approach exploits the numerical sensitivity of a matrix suffering from ill-conditionedness, inheriting the essence from matrix theory (Horn and Johnson 1990). Moreover, our approach retains the intrinsic property of a correlation matrix. Therefore, our approach enjoys both theoretical and numerical merits.

Literature review By far, there is no concensus for the detection of the presence of data collinearity. The presence of data collinearity is often detected through the concepts of the condition number (*CN*), the variance inflation factor (*VIF*) and so forth (Fox and Monette 1992; Curto and Pinto 2007; Kovás et al. 2005). Another criterion is that if there is more than one non-diagonal elements of the correlation matrix having value(s) very close to ±1, the collinearity can be said to be present. All those concepts are built upon the correlation matrix for the explanatory variables in regression models. Each of these existing diagnostic tools has its own advantages and weaknesses. For a thorough discussion, please refer to Belsley (1980).

The concept of *CN* relates to the eigenstructure of the correlation matrix. The *CN* is defined through the ratio of the largest to the smallest eigenvalues of the correlation matrix. A high *CN* value, usually 30 in the literature (Belsley 1980), suggests the presence of data collinearity. Nevertheless, there is still no rule of thumb for the *CN* to reveal the presence of data collinearity. One of the drawbacks of *CN* is that *CN* tends to be inflated, misleading the researchers to believing the presence of data collinearity (Lazaridis 2007). The use of *CN* as the main diagnostic tool should be with caution.

Another widely used diagnostic tool is called the *VIF* that indicates how the variance of the corresponding coefficient is inflated due to data collinearity (Curto and Pinto 2011; Robinson and Schumacher 2009). Naturally, a high *VIF* value for an explanatory variable suggests the presence of data collinearity. Although there is no rule of thumb for *VIF*s, a value of 10 is often adopted, but with caution (O'Brien 2007). Built upon the concept of *VIF*, Fox and Monette (1992) introduced a generalized diagnostic for the collinearity problems. Curto and Pinto (2011) proposed the corrected *VIF*s that incorporate the R^2 values as the adjustments for the original *VIF*s. Lin et al. (2011) proposed a regression method using the concept of *VIF* as the criterion for variable selection. Liao and Valliant (2012) examined the role of *VIF*s in complex survey data. Increases in the use of *VIF* as the major tool for different purposes reveals the increase in its importance in the literature. Based upon this fact, this paper adopts *VIF* as the major tool as a natural choice.

There have been a great many of techniques developed as the remedy for the consequent symptoms resulting from data collinearity. To name a few, the ridge regression (Hoerl and Kennard 1970; McDonald 2009), the LASSO, the bridge regression, the principal component regression (Batah et al. 2009), better estimators, say Liu-type estimators (Liu 2003) or other estimators (Bagheri and Midi 2009), the variable deletion/selection approach (Xin and Zhu 2010), data adjustments (Echambadi and Hess 2007; Shieh 2010) and so forth (Bashtian et al. 2011; Fierro and Bunch 1994; Lin 2008) have been widely used and discussed. Others are referred to Soofi (1990), Leung and Yu (2000) and Næs and Mevik (2001).

In light of the above from the literature, this paper proposes a novel approach, built upon the concept of *VIF*, as a new remedy for data collinearity problems in multiple linear regression. More specifically, we develop a novel optimization model based upon the concept of *VIF* to tackle data collinearity problems. Similar to ridge regression but unlike those intended for variable selection, the proposed *VIF*-based model possesses the feature that all the variables in the regression model are kept, which can be of practical interest (McDonald and Schwing 1973; Schwing and McDonald 1976).

Our contributions This paper addresses the data collinearity problems in multiple linear regression. Our contributions are summarized as follows.

First of all, we propose a novel *VIF*-based optimization model to overcome the data collinearity problems. To the best of our knowledge, there is no such model to date in the literature. The established linearly constrained quadratic programming (LCQP) model is convex, so it is computationally efficient. Moreover, on the one hand, our approach resembles LASSO (Tibshirani 1996) and bridge regression (Frank and Friedman 1993) in the sense that all involve solving associated optimization models to obtain estimators. On the other hand, our approach resembles ridge regression in the sense that the established estimator can be representative of a revised normal equation, to be discussed in the next paragraph.

Second, compared to the ridge regression, our approach tackles the collinearity problems in a more reasonable way. Ridge estimators are derived from imposing a diagonal perturbation matrix, λI, for I an identity matrix of appropriate dimension, on the correlation matrix, viz., $\widehat{\boldsymbol{\beta}}_R = (X^T X + \lambda I)^{-1} X^T \mathbf{y}$, given the centered and scaled data matrix X and $\mathbf{y}$. This can be counterintuitive, for the correlation between a variable and itself is by definition 1, no matter how small the perturbation λ is. One may even find out that the ridge regression can still produce a solution by setting $\lambda = 1$. Our approach preserves the intrinsic of the correlation matrix by imposing a perturbation matrix $\mathcal{W}$ on the correlation matrix, viz., $\widehat{\boldsymbol{\beta}}_{\mathcal{W}} = (X^T X + \mathcal{W})^{-1} X^T \mathbf{y}$, for $\mathcal{W}$ symmetric and having zeros on the diagonal. Such an imposition does not change the values on the diagonal of the original correlation matrix.

Third, our approach provides another tool for regression modelling that evades variable selection when data collinearity problem is present. This can be helpful when certain important variables might be ruled out by the variable selection according to certain criteria, as pointed out in McDonald and Schwing (1973) and Schwing and McDonald (1976). Therefore, our approach accompanies the ridge regression in this aspect. Both the LASSO and the bridge regression are intended for variable selections, so we will not make comparisons therewith.

Lastly, various examples validate our approach. The numerical results indicate that (i) our approach can not only improve the *VIF*s, but the *CN* as well; (ii) statistical significance can be potentially improved; and (iii) estimates with different signs can be corrected.

Organization The rest of the paper is organized as follows. In Sect. 2, we first give an overview of the proposed model, and then specify the proposed LCQP built upon the concept of *VIF*. In Sect. 3, we answer the question of how the superposition of a perturbation matrix affects the OLS estimators. In Sect. 4, we discuss algorithmic and statistical issues. We first visualize certain statistical constraints mathematically. The proposed LCQP can be refined by incorporating the additional statistical constraints. We then state the proposed iterative algorithm for solving the LCQP and the related numerical issues subsequently. In Sect. 5, we mainly focus on the standard t hypothesis testing for the perturbed regression estimators. In Sect. 6, we illustrate various examples that suffer from data collinearity problems. The examples include a sales data drawn from Chatterjee and Hadi (2006) in Sect. 6.1, a famous benchmark diabetes data (Efron et al. 2004) in Sect. 6.2, an application to the collinear dataset (Chatterjee and Hadi 2006) in Sect. 6.3 and one time series data with one variable having wrong sign after being detrended (Shen and Wohlgenant 2010) in Sect. 6.4. Lastly in Sect. 7, we point out certain concluding remarks and the future research directions.

2 Model specification

Overview In this section, we aim at the establishment of the LCQP for addressing data collinearity problems. Before that, we first introduce certain notations that will be used subsequently in the model derivation. We then specify the construction of the model.

The key idea of the proposed approach resides in the superposition of a symmetric non-diagonal perturbation matrix on the correlation matrix. This essentially originates from the matrix theory. When a matrix suffers from ill-posedness, viz., the condition number thereof is very high, a slight perturbation in the data can result in a relatively large change in the eigenstructure of the matrix (Horn and Johnson 1990). Our model inherits the merit of the theory, to show that such perturbations can be obtained through solving a convex LCQP.

We start with the centered and scaled regression model, generally without intercept. In light of Belsley (1980), both the normal equation and the *VIF*s can be considered as a function of the correlation matrix. We employ the first-order Taylor approximation on both the normal equation and the *VIF*s to construct the objective function and the constraints, respectively, for the resultant LCQP. The LCQP has a convex quadratic objective function and a set of linear constraints. The convexity guarantees the computational efficiency in solving the LCQP, as will be shown later in Sect. 6.

2.1 Notations and assumptions

All vectors are written in boldface, and matrices are written in capital letters. We denote by $\mathbf{x} \in \mathbb{R}^m$ a real column vector of dimension m-by-1, with x_i as its ith element, for $i \in \{1, \ldots, m\}$ and $m \in \mathbb{N}$ (set of natural numbers). We write $A \in \mathbb{R}^{m \times n}$ a real matrix of dimension m-by-n, for $m, n \in \mathbb{N}$. We denote a matrix $A \in \mathbb{R}^{m \times n}$ by $\left[a_{i,j}\right]$ in which $a_{i,j}$ is the (i, j)th element of the matrix $A, \forall i \in \{1, \ldots, m\}$ and $\forall j \in \{1, \ldots, n\}$. Occasionally, we use $(A)_{i,j} \equiv a_{i,j}$ as the (i, j)th element of the matrix A. We let $A \equiv \left[a_{i,j}\right]_{i \neq j} \in \mathbb{R}^{m \times m}$ be a squared matrix with all diagonals being zero and with nondiagonals being $a_{i,j}, \forall i \neq j$. A diagonal matrix $A \in \mathbb{R}^{m \times m}$ is denoted by $Diag\left[a_i\right]$ with a_i being the ith diagonal element of A. We denote by $A \succ (\succeq)\, 0$ a positive (semi)definite matrix A. The transpose of a matrix $A \in \mathbb{R}^{m \times n}$ is denoted by $A^T \in \mathbb{R}^{n \times m}$. The inverse of a matrix $A \in \mathbb{R}^{m \times m}$, if it exists, is denoted by $A^{-1} \in \mathbb{R}^{m \times m}$. The determinant of a matrix $A \in \mathbb{R}^{m \times m}$ is denoted by $\det(A)$. The kth submatrix derived from deleting the kth row and column of a matrix A is denoted by $A\{k\}$. The eigen-system of a matrix $A \in \mathbb{R}^{m \times m}$ is denoted by the pair $(\Lambda(A), \mathbf{V}(A))$, where $\Lambda(A) \equiv (\lambda_1, \ldots, \lambda_m)^T$ is the eigenvalues and $\mathbf{V}(A) \equiv (\upsilon_1, \ldots, \upsilon_m)$, with $\upsilon_k \in \mathbb{R}^m$ for $k = 1, \ldots, m$, are the eigenvectors associated.

This paper adopts the standard vector/matrix norms. For a vector $\mathbf{x} \in \mathbb{R}^m$, the standard vector l_2-norm is defined by $\|\mathbf{x}\|_2^2 \equiv \sum_{i=1}^m x_i^2$. For a matrix $A \in \mathbb{R}^{m \times m}$, the standard matrix l_2-norm is defined by $\|A\|_2^2 \equiv \sum_{i,j=1}^m a_{i,j}^2$.

This paper considers a general *centered and scaled* regression model, viz., $\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}$. Let $X \in \mathbb{R}^{n\times k}$ be the centered and scaled data matrix. The matrix product $X^T X$ $(\equiv \varOmega_{XX}) \in \mathbb{R}^{k\times k}$ is the correlation matrix for the independent variables accordingly (Belsley 1984). The correlation coefficient between variable i and variable j is denoted by $\varpi_{i,j}$, viz., $\varOmega_{XX} \equiv \left[\varpi_{i,j}\right]$ with $\varpi_{i,j} = 1$, if $i = j$, and $\varpi_{i,j} \in [-1, 1]$, if $i \neq j$, for $i, j \in \{1, \ldots, k\}$. Let $\mathbf{y} \in \mathbb{R}^n$ be the response (or the dependent) vector, $\boldsymbol{\beta} \in \mathbb{R}^k$ the vector of regression coefficients and $\mathbf{u} \in \mathbb{R}^n$ be the residual vector.

A random variable Y that is normally distributed with mean $\mathbb{E}[Y]$ and variance $Var(Y)$ is denoted by $Y \sim \mathcal{N}(\mathbb{E}[Y], Var(Y))$. If Y is a standard normal random variable, then it is denoted by $Y \sim \mathcal{N}(0, 1)$.

Assumptions to be adopted in this paper are the following. We adopt the traditional linear model assumptions, inclusive of the normality condition for the residuals. We note that the homoskedasticity is required since we shall need the t-statistics as the main hypothesis testing. Moreover, we assume that all the explanatory variables are exogenous.

2.2 Model derivation

We start with the *centered and scaled* regression model

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}. \tag{2.1}$$

The OLS estimators can be secured by the normal equation, as a function of the correlation matrix, in particular

$$\widehat{\boldsymbol{\beta}}(\varOmega_{XX}) = \left(X^T X\right)^{-1} X^T \mathbf{y} \equiv \varOmega_{XX}^{-1}\varOmega_{X\mathbf{y}}, \tag{2.2}$$

where $\varOmega_{X\mathbf{y}} \equiv X^T\mathbf{y}$. The population correlation matrix is by definition symmetric and positive definite. The sample correlation matrix, however, may not be so, depending on the behaviour of the collected non-experimental data, for example Knol and Berge (1989). In this study, we assume that the sample correlation matrix is positive definite, viz., $\varOmega_{XX} \succ 0$. Mathematically, we require the smallest eigenvalue of the correlation matrix to be positive.

In light of Belsley (1980), the *VIF*s of the independent variables can be derived from the diagonal elements of the inverse of the correlation matrix, viz., the diagonals of $\varOmega_{XX}^{-1}$. We denote by $\mathcal{V}_i(\varOmega_{XX})$ the *VIF* for the ith independent variable, in particular

$$\mathcal{V}_i(\varOmega_{XX}) \equiv \left(\varOmega_{XX}^{-1}\right)_{i,i}, \quad \text{for } i \in \{1, \ldots, k\}. \tag{2.3}$$

Our aim is to find a symmetric non-diagonal perturbation matrix $\mathcal{W} \equiv \left[\triangle\varpi_{i,j}\right]_{i\neq j}$, as defined in Sect. 2.1, such that the resulting correlation matrix remains positive definite, viz.,

$$\varOmega_{XX}^{*} = \varOmega_{XX} + \mathcal{W} \succ 0. \tag{2.4}$$

Note that $(\mathcal{W})_{i,j} \equiv \triangle\varpi_{i,j} = 0$, for $i = j$. Statistics imposed with (2.4) are *perturbed*. In our case, the OLS estimators (2.2) imposed with (2.4) become the perturbed OLS estimators, or simply perturbed estimators, in particular

$$\widehat{\boldsymbol{\beta}}_{\mathcal{W}}\left(\Omega^{*}_{XX}\right) \equiv \left(\Omega^{*}_{XX}\right)^{-1}\Omega_{X\mathbf{y}}. \tag{2.5}$$

The *VIF*s (2.3) imposed with (2.4) become the perturbed *VIF*s, in particular

$$\mathcal{V}_i\left(\Omega^{*}_{XX}\right) \equiv \left(\Omega^{*-1}_{XX}\right)_{i,i}, \quad \text{for } i \in \{1, \ldots, k\}. \tag{2.6}$$

We shall employ the perturbation method by applying the first-order Taylor approximation to the perturbed versions of (2.2) and (2.3), viz., (2.5) and (2.6), to establish the objective function and the constraints, respectively, for the LCQP.

Objective function We start with the perturbed normal Eq. (2.5). Applying the first-order Taylor approximation about Ω_{XX} to $\widehat{\boldsymbol{\beta}}\left(\Omega^{*}_{XX}\right)$ suggests

$$\widehat{\boldsymbol{\beta}}_{\mathcal{W}}\left(\Omega^{*}_{XX}\right) = \widehat{\boldsymbol{\beta}}\left(\Omega_{XX}\right) + \sum_{i>j}\widehat{\boldsymbol{\beta}}'_{i,j}\left(\Omega_{XX}\right)\triangle\varpi_{i,j} = \widehat{\boldsymbol{\beta}}\left(\Omega_{XX}\right) + B\boldsymbol{\varpi}, \tag{2.7}$$

where $\boldsymbol{\varpi} \in \mathbb{R}^{n(n-1)/2}$ is the *decision vector* consisting of $\triangle\varpi_{i,j}, \forall i > j$, and the columns of $B \in \mathbb{R}^{n\times(n(n-1)/2)}$ are made of $\widehat{\boldsymbol{\beta}}'_{i,j}\left(\Omega_{XX}\right)$ for each particular pair (i, j), $\forall i \neq j$. We remind the readers of that the perturbation matrix $\mathcal{W} \equiv \left[\triangle\varpi_{i,j}\right]_{i\neq j}$ will be composed of the elements in the decision vector $\boldsymbol{\varpi}$, by rearranging the elements properly. The matrix B can be derived from taking partial derivative of (2.2) with respect to the nondiagonals. In particular,

$$\widehat{\boldsymbol{\beta}}'_{i,j}\left(\Omega_{XX}\right) = \frac{\partial\widehat{\boldsymbol{\beta}}_{i,j}\left(\Omega_{XX}\right)}{\partial\varpi_{i,j}} = -\Omega^{-1}_{XX}\left(\mathbf{e}_i\widehat{\beta}_j + \mathbf{e}_j\widehat{\beta}_i\right), \quad \forall i \neq j$$

wherein $\mathbf{e}_i$ is a column zero vector with 1 at the ith position.

Since we are imposing perturbations on the OLS estimators, the resulting estimators become biased. We define the bias by the difference between the perturbed and the original estimators, viz., $b_{\mathcal{W}} \equiv \widehat{\boldsymbol{\beta}}_{\mathcal{W}}\left(\Omega^{*}_{XX}\right) - \widehat{\boldsymbol{\beta}}\left(\Omega_{XX}\right) = B\boldsymbol{\varpi}$. Therefore, our objective here is to minimize the bias incurred from such perturbations. In particular, we aim to minimize the bias in the regression estimator, viz.,

$$\min_{\boldsymbol{\varpi}\in\mathbb{R}^{n(n-1)/2}} \|B\boldsymbol{\varpi}\|^2 = \boldsymbol{\varpi}^T B^T B\boldsymbol{\varpi}, \tag{2.8}$$

where $\|\cdot\|_2$ is the standard vector l_2-norm. The convexity of (2.8) is obvious, for the product of any matrix A, $A^T A$, is symmetric and positive semidefinite.

VIF constraints We proceed on with the perturbed *VIF*s in (2.6). Applying the first-order Taylor approximation about Ω_{XX} to each $\mathcal{V}_l\left(\Omega_{XX}^*\right)$ suggests

$$\mathcal{V}_l\left(\Omega_{XX}^*\right) = \mathcal{V}_l\left(\Omega_{XX}\right) + \sum_{i>j} \mathcal{V}_l'\left(\Omega_{XX}\right) \triangle\varpi_{i,j}, \quad \text{for } l = 1, \ldots, k. \tag{2.9}$$

The keys to obtain the derivative $\mathcal{V}_l'\left(\Omega_{XX}\right)$ are twofold. First, from matrix theory (Horn and Johnson 1990), we know that

$$\mathcal{V}_l\left(\Omega_{XX}\right) = \det\left(\Omega_{XX}\{l\}\right) / \det\left(\Omega_{XX}\right). \tag{2.10}$$

Second, the determinant of a matrix is the product of its eigenvalues, viz.,

$$\det\left(\Omega_{XX}\right) = \prod_{p=1}^{n} \lambda_p\left(\Omega_{XX}\right) \text{ and } \det\left(\Omega_{XX}\{l\}\right) = \prod_{\hat{p}=1}^{n-1} \lambda_{\hat{p}}\left(\Omega_{XX}\{l\}\right), \tag{2.11}$$

where λ_p and $\lambda_{\hat{p}}$ are the pth and $\hat{p}$th eigenvalues for the matrices Ω_{XX} and $\Omega_{XX}\{l\}$, respectively.

Based upon (2.10) and (2.11), it can be shown that

$$\mathcal{V}_l'\left(\Omega_{XX}\right) = 2\left[\sum_{\hat{q}=1}^{n-1} \frac{\widehat{\upsilon}_{\hat{q},\hat{\imath}}\widehat{\upsilon}_{\hat{q},\hat{\jmath}}}{\lambda_{\hat{q}}\left(\Omega_{XX}\{l\}\right)} - \sum_{q=1}^{n} \frac{\upsilon_{q,i}\upsilon_{q,j}}{\lambda_q\left(\Omega_{XX}\right)}\right] \mathcal{V}_l\left(\Omega_{XX}\right), \tag{2.12}$$

which is a scalar. Note that υ_q and $\widehat{\upsilon}_{\hat{q}}$ are the qth and $\hat{q}$th eigenvectors associated with Ω_{XX} and $\Omega_{XX}\{l\}$, respectively; and $\upsilon_{q,i}$ and $\widehat{\upsilon}_{\hat{q},\hat{\imath}}$ are the ith and $\hat{\imath}$ th elements of the eigenvectors υ_q and $\widehat{\upsilon}_{\hat{q}}$, respectively. We can thereby write (2.9) as

$$\mathcal{V}_l\left(\Omega_{XX}^*\right) = \mathcal{V}_l\left(\Omega_{XX}\right) + \mathbf{v}_l^T\boldsymbol{\varpi}, \quad \text{for } l = 1, \ldots, k, \tag{2.13}$$

for $i \neq j$ and $\hat{\imath} \neq \hat{\jmath}$ in (2.12). Note that the vector $\mathbf{v}_l$ in (2.13) is composed of (2.12) for different pairs (i, j).

From (2.13) we define the difference between the perturbed and the original *VIF*s by

$$d_l\left(\boldsymbol{\varpi}\right) \equiv \mathcal{V}_l\left(\Omega_{XX}^*\right) - \mathcal{V}_l\left(\Omega_{XX}\right) = \mathbf{v}_l^T\boldsymbol{\varpi}, \quad \text{for } l = 1, \ldots, k.$$

Our aim is to decrease the *VIF*s by imposing a perturbation on the correlation matrix, viz., we want $d_l\left(\boldsymbol{\varpi}\right) < 0$. To this end, according to the authors' experience, it suffices to consider only the independent variable with the largest *VIF* value, in particular, we consider

$$\mathcal{V}_{k^*}\left(\Omega_{XX}^*\right) = \mathcal{V}_{k^*}\left(\Omega_{XX}\right) + \mathbf{v}_{k^*}^T\varpi$$

where $k^* = \max_{1\leq j\leq k}\left\{j|\mathcal{V}_j\left(\Omega_{XX}\right) \geq \mathcal{V}_i\left(\Omega_{XX}\right), \text{ for } i \neq j\right\}$. We denote such a constraint as

$$\mathcal{V}_{\max}\left(\Omega_{XX}^*\right) = \mathcal{V}_{\max}\left(\Omega_{XX}\right) + \mathbf{v}_{\max}^T\boldsymbol{\varpi} \text{ so that } d_{\max}\left(\boldsymbol{\varpi}\right) < 0 \tag{2.14}$$

Note that (2.14) constitutes a linear constraint, and thus convex. We are now ready to state the LCQP.

A quadratic programming model Combining (2.8) and (2.14), we form the following norm-minimization model

$$\begin{aligned} &\min \boldsymbol{\varpi}^T B^T B \boldsymbol{\varpi} + \rho \boldsymbol{\varpi}^T \boldsymbol{\varpi} \\ &\text{s.t.} - \mathbf{v}_{\max}^T \boldsymbol{\varpi} = v_r \text{ and } \boldsymbol{\varpi} \in \mathbb{R}^{n(n-1)/2} \end{aligned} \tag{2.15}$$

where $\rho(>0)$ is a trade-off parameter and $v_r(>0)$ is the reduction in the *VIF*. An intuitive interpretation of (2.15) is the following. We aim to find a perturbation matrix (symmetric and zeros on diagonal) so that the largest *VIF* is reduced, meanwhile minimizing the bias resulting from the perturbation.

The strict convexity of the objective function guarantees that the global minimizer to (2.15) can be found, in light of the K–K–T optimality condition (Bazaraa et al. 2006) which in our case is both necessary and sufficient. In particular, the optimal perturbation vector $\boldsymbol{\varpi}^*$ is given by

$$\boldsymbol{\varpi}^* = \frac{-v_r}{\mathbf{v}_{\max}^T \left(B^T B + \rho I\right)^{-1} \mathbf{v}_{\max}} \left(B^T B + \rho I\right)^{-1} \mathbf{v}_{\max}. \tag{2.16}$$

The trade-off parameter ρ plays two roles in the model. First, we wish the norm of perturbation $\boldsymbol{\varpi}^T \boldsymbol{\varpi}$ is small enough so that the model works well. If the norm were too large, the model could have poorly performed. The trade-off parameter ρ serves as the penalty imposed on the norm of $\boldsymbol{\varpi}^T \boldsymbol{\varpi}$. Second, the addition of the trade-off parameter ensures that the objective function is well-posed, viz., $B^T B + Diag\,[\rho] \succ 0$, for $\rho > 0$. The choices of the parameters ρ and v_r will be discussed more in Sect. 4.

3 On perturbed estimators

In this section, we examine how the superposition of a perturbation matrix affects the OLS estimators asymptotically. The key to analysis originates from Miller (1981) who gave the inversion results of a sum of several matrices. The main result on the inversion of a sum of two matrices is given by

$$(A + B)^{-1} = A^{-1} - \left(I + A^{-1} B\right)^{-1} A^{-1} B A^{-1}, \tag{3.1}$$

which is crucial to the following analysis. The proof of (3.1) is simply by multiplying $(A + B)$ on both sides.

Recall that our approach involves imposing a symmetric non-diagonal perturbation matrix on the correlation matrix, viz., (2.4). It is obvious that the perturbed estimators are biased. We characterize the bias in terms of the perturbation matrix in what follows.

Expectation and variance Starting with (2.5) together with (3.1), it follows that

$$\widehat{\boldsymbol{\beta}}_{\mathcal{W}}\left(\Omega^*_{XX}\right)=\widehat{\boldsymbol{\beta}}_{OLS}-\left[\left(\Omega^*_{XX}\right)^{-1}\mathcal{W}\right]\widehat{\boldsymbol{\beta}}_{OLS}. \tag{3.2}$$

which is a function of the OLS estimators $\widehat{\boldsymbol{\beta}}_{OLS}$. It follows that the difference between the original and the perturbed OLS estimators us given by

$$\widehat{\boldsymbol{\beta}}_{OLS}-\widehat{\boldsymbol{\beta}}_{\mathcal{W}}=\left[\left(\Omega^*_{XX}\right)^{-1}\mathcal{W}\right]\widehat{\boldsymbol{\beta}}_{OLS}. \tag{3.3}$$

Taking expectation on both sides suggests

$$\boldsymbol{\beta}-\mathbb{E}\left[\widehat{\boldsymbol{\beta}}_{\mathcal{W}}\right]=\left[\left(\Omega^*_{XX}\right)^{-1}\mathcal{W}\right]\boldsymbol{\beta}, \tag{3.4}$$

where $\mathbb{E}\left[\widehat{\boldsymbol{\beta}}_{OLS}\right]\equiv\boldsymbol{\beta}$, the unknown population parameters. The identity (3.4) indicates that the bias between the population parameters and the expected perturbed estimators is $\left[\left(\Omega^*_{XX}\right)^{-1}\mathcal{W}\right]\boldsymbol{\beta}$. Thus, if the perturbation matrix $\mathcal{W}$ is appropriately controlled, so is the bias.

The covariance of the established estimators $\widehat{\boldsymbol{\beta}}_{\mathcal{W}}$ can also constructed from (3.2), which suggests

$$Cov\left(\widehat{\boldsymbol{\beta}}_{\mathcal{W}}\right)=\sigma^2 Q\left(\mathcal{W}\right), \tag{3.5}$$

where $Q\left(\mathcal{W}\right)\equiv\left[I-\left(\Omega^*_{XX}\right)^{-1}\mathcal{W}\right]\left[I-\left(\Omega^*_{XX}\right)^{-1}\mathcal{W}\right]^T$ is a function of the perturbation matrix $\mathcal{W}$.

Bounded bias We now show that the bias (3.3) is effectively bounded above. From (3.3), it is not hard to see that the length of the bias in terms of usual vectoral l_2-norms becomes

$$\left\|\widehat{\boldsymbol{\beta}}_{\mathcal{W}}-\widehat{\boldsymbol{\beta}}_{OLS}\right\|_2\leq\left(\left\|\left(\Omega^*_{XX}\right)^{-1}\right\|_2\left\|\widehat{\boldsymbol{\beta}}_{OLS}\right\|_2\right)\|\mathcal{W}\|_2\equiv k\,\|\mathcal{W}\|_2, \tag{3.6}$$

wherein $k\equiv\left\|\left(\Omega^*_{XX}\right)^{-1}\right\|_2\left\|\widehat{\boldsymbol{\beta}}_{OLS}\right\|_2$. Note that the norms for the matrices $\left(\Omega^*_{XX}\right)^{-1}$ and $\mathcal{W}$ in (3.6) are standard matrix l_2-norm, while that for the vector $\widehat{\boldsymbol{\beta}}_{OLS}$ is vector l_2-norm, as defined in Sect. 2.1. This indicates that the bias resulting from the superposition of a perturbation matrix is controlled by the norm of the perturbation matrix. Since $\mathcal{W}$ is composed of the elements of the perturbation vector $\boldsymbol{\varpi}$, the relationship (3.6) can be further represented as

$$\left\|\widehat{\boldsymbol{\beta}}_{\mathcal{W}}-\widehat{\boldsymbol{\beta}}_{OLS}\right\|_2\leq k\,\|\mathcal{W}\|_2=k'\,\|\boldsymbol{\varpi}\|_2, \tag{3.7}$$

where $k'=2k$. From (3.7), it becomes obvious that, as long as the lengths of the perturbations generated by the LCQP (2.15) are minimized, the biases in the estimators are thereby minimized.

4 Algorithm

Overview In this section, our aim is twofold. First, we bring up certain statistical issues, mainly on the perturbation matrix as well as the R^2 statistic in Sect. 4.1. Mathematically, we visualize them as the statistical constraints, as potential side constraints for the LCQP (2.15). Second, we state the proposed *VIF*-based algorithm, the settings for the model inputs and certain numerical issues in Sect. 4.2.

4.1 Statistical issues

There are certain issues for the perturbation matrix derived from (2.15) as well as the goodness of fit for the perturbed model. We discuss the issues in what follows.

Approximate confidence bounds Let $\Omega^*_{XX} \equiv \left[\varpi^*_{i,j}\right]$. We require that the perturbed correlation coefficients should lie within the confidence interval under the null hypothesis $H_0 : \left\{\varpi^*_{i,j} = \varpi_{i,j}\right\}$ for a specific level of confidence. Let the null hypothesis be $H_0 : \left\{\varpi^*_{i,j} = \varpi_{i,j}\right\}$ against the alternative $H_1 : \left\{\varpi^*_{i,j} \neq \varpi_{i,j}\right\}$. We adopt the Fisher's Z-transform on the correlation coefficients, given by

$$Z_{\boldsymbol{\varpi}^*} = \frac{1}{2}\ln\left(\frac{1+\varpi^*_{i,j}}{1-\varpi^*_{i,j}}\right), \quad \text{for } i \neq j,$$

following approximately an normal distribution with mean $Z_{\boldsymbol{\varpi}}$ and variance $\sigma^2_Z = \frac{1}{n-3}$, viz., $Z_{\boldsymbol{\varpi}^*} \sim \mathcal{N}\left(Z_{\boldsymbol{\varpi}}, (n-3)^{-1}\right)$. The Fisher's Z-statistic ($\mathcal{Z}$) is defined as

$$\mathcal{Z} \equiv \frac{(Z_{\boldsymbol{\varpi}^*} - Z_{\varpi})}{\sigma_Z} = \frac{\sqrt{n-3}}{2}\ln\left[\frac{\left(1+\varpi^*_{i,j}\right)\left(1-\varpi_{i,j}\right)}{\left(1-\varpi^*_{i,j}\right)\left(1+\varpi_{i,j}\right)}\right] \to \mathcal{N}(0,1).$$

Since both the standard deviation σ_Z and $\varpi_{i,j}, \forall i, j \in \{1, \ldots, k\}$, are known, we can derive the confidence bounds for $\varpi^*_{i,j}$ accordingly. It is not hard to see that the following constraint satisfies the $100 \times (1-\alpha)$ % confidence interval for $\varpi^*_{i,j}$

$$\frac{\frac{(1+\varpi_{i,j})}{(1-\varpi_{i,j})}\exp\left(\frac{2z_{\frac{\alpha}{2}}}{\sqrt{n-3}}\right)-1}{\frac{(1+\varpi_{i,j})}{(1-\varpi_{i,j})}\exp\left(\frac{2z_{\frac{\alpha}{2}}}{\sqrt{n-3}}\right)+1} \leq \varpi^*_{i,j} \leq \frac{\frac{(1+\varpi_{i,j})}{(1-\varpi_{i,j})}\exp\left(\frac{2z_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}\right)-1}{\frac{(1+\varpi_{i,j})}{(1-\varpi_{i,j})}\exp\left(\frac{2z_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}\right)+1}, \quad \text{for } i \neq j \tag{4.1}$$

Remark 4.1 (Approximate bounds) Unfortunately, the confidence bounds (4.1) serves merely as an approximation for the perturbed correlation coefficient $\varpi^*_{i,j}$, for $i \neq j$. The use of the Fisher's Z-transform relies on the normality condition. The Fisher's Z-statistic can perform very poorly if the sample size is not large enough, say $n < 500$

(Paul 1989). Moreover, there is so far no well-founded hypothesis testing procedure for testing the relationship between the perturbed and the original correlation coefficients, viz., $H_0 : \left\{\varpi^*_{i,j} = \varpi_{i,j}\right\}$ against the alternative $H_1 : \left\{\varpi^*_{i,j} \neq \varpi_{i,j}\right\}$. We bring up (4.1) for purpose of completion.

Range of correlation coefficients We require that the perturbed correlation matrix still retains its own characteristic, viz., every correlation coefficient should be within the -1 to 1 range. Mathematically, we have

$$\varpi^*_{i,j} \in (-1, 1), \quad \text{for } i \neq j, \text{and } \varpi^*_{i,j} = 1, \quad \text{for } i = j. \tag{4.2}$$

Sign restrictions The signs of the perturbation should satisfy some rules, if any. That is, for some specific (i, j)th element, the corresponding perturbation is restricted in sign,

$$\triangle\varpi_{i,j} = 0 \text{ or } \pm\triangle\varpi_{i,j} > 0, \text{ for } (i, j) \in \mathcal{I} \tag{4.3}$$

where $\mathcal{I}$ is an index set subject to sign restrictions.

Goodness of fit The value of R^2 must be less than unity. The R^2 statistic is defined by $R^2 = 1 - C_1 (\widetilde{\mathbf{y}} - \widehat{\mathbf{y}})^T (\widetilde{\mathbf{y}} - \widehat{\mathbf{y}})$, where $\widetilde{\mathbf{y}} \equiv (\mathbf{y} - \bar{\mathbf{y}})$, $\widehat{\mathbf{y}} = X\widehat{\boldsymbol{\beta}}$ and $C_1 \equiv \left(\widetilde{\mathbf{y}}^T\widetilde{\mathbf{y}}\right)^{-1}$. It is obvious that R^2 is a function of the regression estimator $\widehat{\boldsymbol{\beta}}$, and therefore a function of the correlation matrix. In particular, letting $R^2 \equiv \mathcal{R}(\varOmega_{XX})$, we see that

$$\mathcal{R}(\varOmega_{XX}) = 1 - C_1 \left(\widetilde{\mathbf{y}} - X\widehat{\boldsymbol{\beta}}(\varOmega_{XX})\right)^T \left(\widetilde{\mathbf{y}} - X\widehat{\boldsymbol{\beta}}(\varOmega_{XX})\right).$$

Imposing a perturbation matrix $\mathcal{W}$ such that (2.4) holds, the first-order Taylor approximation of $\mathcal{R}\left(\varOmega^*_{XX}\right)$ about $\varOmega_{XX}$ suggests

$$\mathcal{R}\left(\varOmega^*_{XX}\right) = \mathcal{R}(\varOmega_{XX}) + \sum_{i>j} \mathcal{R}'_{i,j}(\varOmega_{XX}) \triangle\varpi_{i,j}, \tag{4.4}$$

in which $\mathcal{R}'(\varOmega_{XX})$ is the partial derivative with respect to $\varpi_{i,j}$, $\forall i > j$. It is not hard to see that the $\mathcal{R}'_{i,j}(\varOmega_{XX})$ in (4.4) is given by, for $i \neq j$,

$$\mathcal{R}'_{i,j}(\varOmega_{XX}) = 2C_1 \left(\widehat{\boldsymbol{\beta}}^T - \widetilde{\mathbf{y}}^T X \varOmega^{-1}_{XX}\right) \left(\mathbf{e}_i\widehat{\beta}_j + \mathbf{e}_j\widehat{\beta}_i\right), \tag{4.5}$$

which is a scalar and $\mathbf{e}_i$ is a column zero vector with 1 at the ith position. Expressed in the vector/matrix notation, (4.5) becomes $\mathcal{R}\left(\varOmega^*_{XX}\right) = \mathcal{R}(\varOmega_{XX}) + \mathbf{r}^T\varpi$. Similarly, we want the difference between the perturbed and original R^2 to remain non-negative, viz.,

$$\mathcal{R}\left(\varOmega^*_{XX}\right) - \mathcal{R}(\varOmega_{XX}) = \mathbf{r}^T\boldsymbol{\varpi} \geq 0. \tag{4.6}$$

Remarkably, the R^2 constraint (4.6) can be helpful when we lose grip on the R^2 for the perturbed model.

Refined LCQP model The original LCQP (2.15) can be refined by incorporating the side constraints (4.1)–(4.3) and (4.6), viz.,

$$\min_{\boldsymbol{\varpi} \in \mathbb{R}^{n(n-1)/2}} \left\{ \boldsymbol{\varpi}^T \left(B^T B + \rho I \right) \boldsymbol{\varpi} : -\mathbf{v}_{\max}^T \boldsymbol{\varpi} = v_r, (4.1)-(4.3) \text{ and } (4.6) \right\}. \quad (4.7)$$

In this paper, the statistical constraints (4.1)–(4.3) and (4.6) will be treated as posterior test.

On positive definiteness of perturbed correlation matrix We now return to (2.4) left assertive. In general, $\mathcal{W}$ is indefinite in essence, so there is no guarantee that (2.4) is true. Computationally, however, if one can ensure that the minimal eigenvalue of Ω_{XX} is greater than or equal to the maximal eigenvalue of $\mathcal{W}$, (2.4) will hold. Specifically, assuming that Ω_{XX}^* is positive definite, then, for any $\mathbf{x} \neq 0$, we see that

$$\mathbf{x}^T \Omega_{XX}^* \mathbf{x} = \mathbf{x}^T \left(\Omega_{XX} + \mathcal{W} \right) \mathbf{x} > \left(\min_{1 \le i \le n} \{\lambda_i (\Omega_{XX})\} - \max_{1 \le i \le n} \{\lambda_i (\mathcal{W})\} \right) \mathbf{x}^T \mathbf{x} > 0,$$

which implies $\min_{1 \le i \le n} \{\lambda_i (\Omega_{XX})\} \ge \max_{1 \le i \le n} \{\lambda_i (\mathcal{W})\}$. More numerical evidences will be in Sect. 6.

4.2 Proposed algorithm, parameter settings and numerical issues

Before introducing the proposed algorithm, we bring up certain numerical issues that relate to the design of the proposed algorithm. The proposed algorithm requires certain exogeneous inputs to begin with. We shall describe the model input and the settings thereof. We state the general structure of the proposed algorithm thereafter.

4.2.1 Numerical issues and algorithm inputs

Recall that the proposed LCQP (2.15) tackles the data collinearity problem by reducing the largest *VIF* while keeping the resultant biases well controlled. The hope to find a perturbation matrix that can reduce the largest *VIF* down to a specified level in a one-step fashion can be of greed and is risky rendering the LCQP (2.15) to collapse. Compared to the ridge regression, designing an iterative algorithm for our approach is of necessity, as the inevitable trade-off for finding a more general symmetric non-diagonal perturbation matrix. We shall show that the computational efficiency of the designed algorithm for the LCQP (2.15) does not frustrate us much, albeit iterative.

Parameters for the VIF-based model For the LCQP (2.15), we need to specify the trade-off parameter ρ and the reduction in *VIF* v_r. The trade-off parameter ρ is by default set to 5. We will show that the choice of ρ does not affect much the performance of (2.15).

The choice of v_r is more important than that for ρ, because the choice of v_r relates directly to the performance (or quality) of the solution to the LCQP model. A moderate level for v_r, say 5, is oftentimes a good choice. Any value greater than 10 would not be recommended. This paper chooses $v_r = 1$ by default. A more elaborate setting for v_r is the dynamic adjustment. Keeping v_r at a constant level may limit the performance of our approach, especially at few steps before termination. A recommended setting for v_r would be in decreasing order, viz., $v_r^i > v_r^{i+1} > 0$ for i the iteration number. The only trade-off for the dynamic adjustment is, however, that the number of iterations increases. This can be offset by the computational efficiency in solving (2.15).

Parameters for the algorithm Since the proposed algorithm is iterative in nature, we need to specify the initial values for the algorithm to begin with. The OLS estimators $\widehat{\boldsymbol{\beta}}_{OLS}$ for (2.1) are chosen as the initial estimators for the algorithm. The correlation matrix Ω_{XX} as well as the *VIF*s $\left\{\mathcal{V}_j\left(\Omega_{XX}\right), j = 1, \ldots, k\right\}$ are known from the data beforehand.

We now specify the stopping criterion for our iterative algorithm. Recall that what LCQP (2.15) does is to reduce the largest *VIF* while keeping the bias minimized. And since our aim is to mitigate the harm from the presence of data collinearity, it is reasonable to set an appropriately desired level for the largest *VIF* to go to. We denote the desired level by v_l. For the purpose of comparison, we shall let v_l be 3 by default, depending on the examples.

Algorithm inputs By and large, the algorithm inputs can be represented as a six-tuple vector $\mathcal{P}$, in particular

$$\mathcal{P} = \left(\Omega_{XX}, \left\{\mathcal{V}_j\left(\Omega_{XX}\right), j = 1, \ldots, k\right\}, \widehat{\boldsymbol{\beta}}_{OLS}, \rho, v_r, v_l\right) \tag{4.8}$$

4.2.2 A VIF-based algorithm

We now present our proposed algorithm. A general algorithmic structure for the algorithm is given below.

```
**WHILE** $\max_{j=1,\ldots,k}\left\{\mathcal{V}_j\left(\Omega_{XX}\right)\right\} > v_l$
    **Step 1. (Initialization)** Input $\mathcal{P}$ as in (4.8)
    **Step 2. (Construct matrices)** Determine the matrices $B$ and $\mathbf{v}_{\max}$
    **Step 3. (Solve for optimality)** Solve (2.15) or use (2.16) to obtain $\varpi^*$
    **Step 4. (Posterior Tests)** Test if the solution from **Step 3** satisfies (4.1)–(4.3)
    **Step 5. (Update)** Update $\Omega_{XX} \to \Omega_{XX} + \mathcal{W}^*$, $\widehat{\beta}\left(\Omega_{XX}\right) \to \widehat{\beta}\left(\Omega_{XX}\right) + B\varpi^*$ and $\mathcal{V}_{\max}\left(\Omega_{XX}\right) \to$
        $\mathcal{V}_{\max}\left(\Omega_{XX}\right) - \mathbf{v}_{\max}^T\varpi^*$.
ENDWHILE
```

A few remarks on the *VIF*-based algorithm are made. First of all, the use of the refined LCQP (4.7) incorporates the additional statistical constraints (4.1)–(4.3) and (4.6), as given in Sect. 4.1. However, as numerical results suggest in Sect. 6, there is no need to impose additional difficulty on the LCQP (2.15). We treat the statistical constraints (4.1)–(4.3) and (4.6), if not used, as posterior tests.

Regarding the **Step 4** of the proposed algorithm, the readers must be aware of the fact that the superposition of the perturbation matrix $\mathcal{W}^*$ has a universal effect on all *VIF*s (and, of course, the estimators). Recall that the constraint of the LCQP (2.15) is built upon the independent variable with the largest *VIF*, and the index thereof is determined by $k^* = \max_{1 \leq j \leq k} \left\{ j | \mathcal{V}_j \left(\Omega_{XX} \right) \geq \mathcal{V}_i \left(\Omega_{XX} \right), i \neq j \right\}$. The perturbation matrix $\mathcal{W}^*$ is, however, imposed upon each and every non-diagonal element of the correlation matrix. There is no escape that all *VIF*s change due to the superposition. Fortunately, the changes in the *VIF*s are beneficial in the sense that only those pathological *VIF*s are reduced while those good *VIF*s fluctuates slightly around their original values. This will later be clear in Sect. 6. By and large, numerically, we effectively calculate all *VIF*s in **Step 4** and the index in the constraint $-\mathbf{v}_{\max}^T \boldsymbol{\varpi} = v_r$ (or $\mathbf{r}^T \boldsymbol{\varpi} \geq 0$) varies over iterations.

Maximal number of iterations The objective of this study is to mitigate the effect of the presence of the data collinearity, instead of the annihilation thereof. As a matter of fact, there is no way to eliminate the data collinearity, even if it is not severe at all, given we are working with non-experimental data. Hence, it is not our goal to reduce the VIFs down to level 0. Setting an appropriate desired level at which the *VIF*s go to, the maximal number of iteration must be finite. More precisely, given a desired level v_l and the reduction in *VIF* v_r, the maximal total number of iterations required by the algorithm is

$$M = \left\lceil \frac{\mathcal{V}_{\max} \left(\Omega_{XX} \right) - v_l}{v_r} \right\rceil < \infty,$$

wherein all quantities are finite and $\lceil a \rceil$ is the least integer great than or equal to $a \in \mathbb{R}$. As a consequence, it follows that

$$\widetilde{k}_M = k' \left\lceil \frac{\mathcal{V}_{\max} \left(\Omega_{XX} \right) - v_l}{v_r} \right\rceil \left(\sup_{0 \leq i \leq M} \left\| \boldsymbol{\varpi}^i \right\|_2 \right) < \infty,$$

for k' in (3.7).

4.3 Perturbed estimators revisited

Having described the proposed VIF-based algorithm, we examine how the iterative structure affects the bias of our regression estimators $\widehat{\boldsymbol{\beta}}_{\mathcal{W}} \left(\Omega_{XX}^* \right)$.

We return to (3.7). Let i be the iteration number and M be the maximal iteration number at which the algorithm terminates. The inequality (3.7) becomes

$$\left\| \widehat{\boldsymbol{\beta}}_{\mathcal{W}}^i \left(\Omega_{XX}^* \right) - \widehat{\boldsymbol{\beta}} \left(\Omega_{XX} \right) \right\|_2 \leq k' \left\| \boldsymbol{\varpi}^i \right\|_2, \quad \text{for } i = 1, \ldots, M$$

wherein $\widehat{\boldsymbol{\beta}}_{\mathcal{W}}^i \left(\Omega_{XX}^* \right)$ and $\boldsymbol{\varpi}^i$ are the results at the i th iteration of the algorithm. Hence, upon termination, the overall bias generated by the algorithm is

$$\sum_{i=1}^{M} \left\| \widehat{\boldsymbol{\beta}}_{\mathcal{W}}^i \left(\Omega_{XX}^* \right) - \widehat{\boldsymbol{\beta}} \left(\Omega_{XX} \right) \right\|_2 \leq \sum_{i=1}^{M} k' \left\| \boldsymbol{\varpi}^i \right\|_2 \leq \widetilde{k}_M, \tag{4.9}$$

where $\widetilde{k}_M \equiv k'M\sup_{0\leq i\leq M}\left\|\boldsymbol{\varpi}^i\right\|_2$ is a constant. Note that the supremum operator can be replaced by the maximum operator, as long as M is finite. Since $\boldsymbol{\varpi}^i$ constitutes the global minimizer of the LCQP (2.15) at iteration i, the quantity $\sup_{0\leq i\leq M}\left\|\boldsymbol{\varpi}^i\right\|_2$ is well controlled and so is the overall bias. Based upon the fact above, we term the resulting estimator $\widehat{\boldsymbol{\beta}}^M_{\mathcal{W}}\left(\Omega^*_{XX}\right)$ the *least–accumulative–bias estimators* (LABs).

5 Hypothesis testing

In this section, we discuss the hypothesis testing on the perturbed estimators. As the objective function of (2.15) suggests, the superposition of the perturbation on the correlation matrix results in changes in the regression estimators. It is necessary to perform the standard t test on the perturbed regression estimators. We also show that the the t-statistic is effectively a function of the *VIF*s.

We implement the standard t test on the null hypothesis $H_0 : \left\{\beta_j\left(\Omega^*_{XX}\right)=0\right\}$ against $H_1 : \left\{\beta_j\left(\Omega^*_{XX}\right)\neq 0\right\}$. The purpose is to ensure that the perturbed regression estimators still retain their statistical significance.

The t-statistic for the jth estimator is defined by $t_j=\widehat{\beta}_j/se\left(\hat{\beta}_j\right)$, for $j=1,\ldots,k$. The standard errors of the regression estimators are effectively functions of the *VIF*s, in particular,

$$se\left(\widehat{\beta}_j\right)=\sqrt{\sigma^2\mathcal{V}_j\left(\Omega^*_{XX}\right)}\text{ with }\sigma^2=\left\|\left(\mathbf{y}-\bar{\mathbf{y}}-X\widehat{\boldsymbol{\beta}}\left(\Omega^*_{XX}\right)\right)\right\|_2^2,$$

for a centered and scaled model (2.1). The perturbed t-statistic therefore is defined as

$$t_j\left(\Omega^*_{XX}\right)\equiv\widehat{\boldsymbol{\beta}}\left(\Omega^*_{XX}\right)/\sqrt{\sigma^2\mathcal{V}_j\left(\Omega^*_{XX}\right)}. \tag{5.1}$$

From the relationship between $t_j\left(\Omega^*_{XX}\right)$ and $\mathcal{V}_j\left(\Omega^*_{XX}\right)$ in (5.1), it is obvious that the reductions in *VIF*s increase the values of the t-statistics.

6 Numerical examples

We apply the LCQP (2.15) and the proposed algorithm in Sect. 4.2 to four examples of two data types. The first two examples are of cross-section. One of them is the diabetes dataset discussed in Efron et al. (2004), and the other is a collinear dataset (Chatterjee and Hadi 2006) in that the correlation coefficients among independent variables are all greater than 0.9. The rest are time-series datasets. One of them is drawn from Chatterjee and Hadi (2006), and the other is from Shen and Wohlgenant (2010). All experiments are implemented in MATLAB R2010b with Pentium 4, 3 GHz CPU and 1G RAM.

Table 1 The summary of the original information for the Diabete data, derived from the ordinary least squares

Vars	*age*	*sex*	*bmi*	*map*	*tc*	*CN*
$\widehat{\beta}_{OLS}$	−0.48	−11.42	24.76	15.45	−37.72	470.08
*VIF*s	**1.22**	**1.28**	**1.51**	**1.46**	**59.2**	–
SE	59	60.46	65.7	64.6	411.46	–
Vars	*ldl*	*hdl*	*tch*	*ltg*	*glu*	R^2
$\widehat{\beta}_{OLS}$	22.7	4.81	8.43	35.78	3.22	0.52
*VIF*s	**39.19**	**15.4**	**8.89**	**10.08**	**1.49**	–
SE	334.79	209.87	159.45	169.75	65.16	–

Significance of bold shows the reduction in VIFs

Table 2 The summary of the original information for the Diabete data, derived from the proposed algorithm

Vars	*age*	*sex*	*bmi*	*map*	*tc*	*CN*
$\widehat{\beta}_{\mathcal{W}}$	−0.2	−11.31	25.56	15.99	−37.12	20.46
*VIF*s	**1.17**	**1.32**	**2.26**	**1.52**	**2.99**	Iters
SE	57.88	61.52	80.59	65.99	92.76	93
Vars	*ldl*	*hdl*	*tch*	*ltg*	*glu*	R^2
$\widehat{\beta}_{\mathcal{W}}$	22.92	4.28	9.01	36.45	3.76	0.52
*VIF*s	**2.09**	**1.75**	**2.18**	**2.18**	**1.4**	CPU
SE	77.45	70.81	79.1	79.19	63.46	6.48

Significance of bold shows the reduction in VIFs

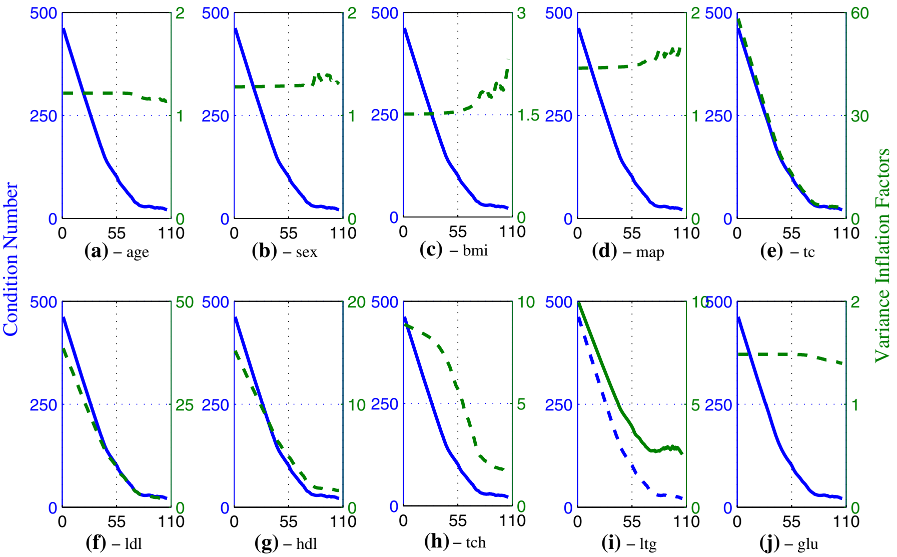
6.1 The diabetes dataset

The diabetes data, as discussed in Efron et al. (2004), has 10 explanatory variables, inclusive of *age*, *sex*, *bmi*, *map*, *tc*, *ldl*, *hdl*, *tch*, *ltg* and *glu*; and there is one response variable. The context of the dataset is to construct a model to examine the relationship among those variables. The basic regression information is summarized in Table 1, consisting of the OLS estimators, *VIF* values and the standard error.

Applying the LCQP (2.15) using the algorithm to the model yields the perturbed information in Table 2. We see that the R^2 value for the perturbed model remains at the same level as that from the OLS results. The *CN* has dropped down enormously from 470 to 21.

Figure 1 indicates the relationship between the *CN* and the *VIF*s. From the figure, we first observe that the *VIF* s for (a) *age*, (b) *sex*, (c) *bmi*, (d) *map* and (j) *glu* basically fluctuate around their original values, even though the *VIF*s for (c) *bmi* and (d) *map* increase slightly. As for those variables with pathological *VIF*s, their *VIF*s suggest a tendency to decrease over the iterations.

Figure 2 concerns two things. First, Fig. 2a shows the relationship between the *CN* and the perturbation norm $\|\boldsymbol{\varpi}\|_2$ over the iterations. By and large, the norm increases



(a) – age **(b)** – sex **(c)** – bmi **(d)** – map **(e)** – tc

(f) – ldl **(g)** – hdl **(h)** – tch **(i)** – ltg **(j)** – glu

Fig. 1 **a–j** Represent the relationship between the condition number and the *VIF*s for each explanatory variable in the diabetes dataset. (*Solid line*—left axis; *dashed line*—right axis)

as the *CN* decreases, even though there are certain fluctuations after the 60th iteration. This fact indicates that it indeed takes more effort for the perturbation to reduce to *VIF*s, as the ill-posedness is mitigated. The basically reflects the matrix property (Horn and Johnson 1990). Second, Fig. 2b gives the numerical validation of the postive definiteness assumption (2.4). As in Sect. 4.1, we show that, as long as the minimal eigenvalue of Ω_{XX} is greater than the maximal eigenvalue of $\mathcal{W}$, the positive definiteness of Ω^*_{XX} remains.

6.2 An application to collinear data

We make an attempt to deal with collinear dataset as an example. The context of the collinear dataset results from a study on the equal opportunity in public education in United States. The objective was to examine the effect of school inputs on students' achievements. More details are referred to Chatterjee and Hadi (2006) and the related literature therein. There are three independent variables (*FAM*, *PEER* and *SCHOOL*), 1 dependent variable (*ACHV*) and 70 random measurements.

The regression information is summarized in Table 3 below. As the authors mentioned, a high F-statistic value (5.72) indicates that the three variables are valid as the explanatory variables, although the t-statistics reveal statistical insignificance individually.

The information in Table 3, as well as the correlation matrix below, all suggest the strong collinearity among independent variables. The strong linear structure between pairs of the three variables does affect the estimators obtained in Table 3.

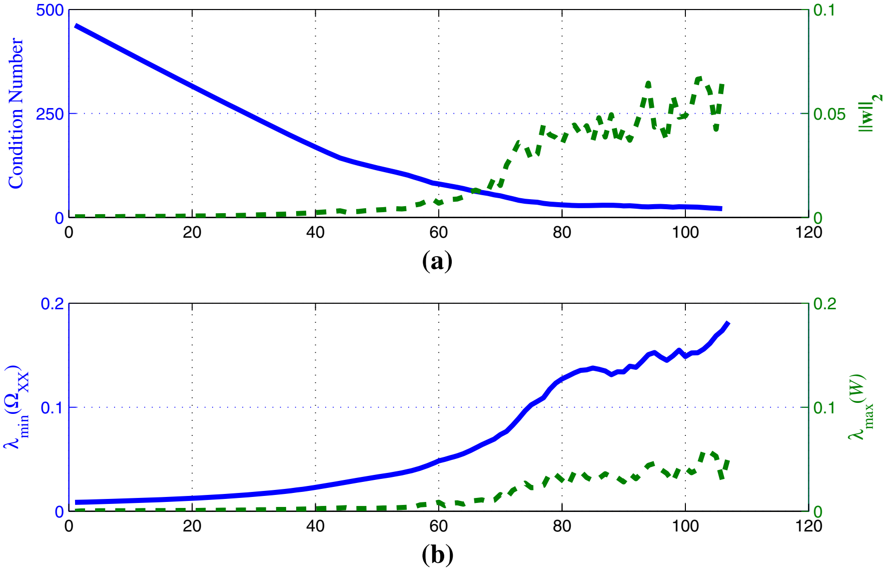


Fig. 2 **a** Relationship between the condition number and the perturbation norm; **b** concerns the positive definiteness property of the perturbed correlation matrix for diabetes data. (*Solid line*—left axis; *dashed line*—right axis)

Table 3 The summary of the original information for the education data, derived from the ordinary least squares

Variables	*FAM*	*PEER*	*SCHOOL*	*CN*	R^2
$\widehat{\beta}_{OLS}$	1.16	1.74	−1.91	393.97	0.19
*VIF*s	**38.443**	**31.478**	**88.372**	–	–
SE	12.52	11.33	18.98	–	–

Significance of bold shows the reduction in VIFs

Table 4 The correlation matrix for variables FAM, PEER and SCHOOL

Variables	*FAM*	*PEER*	*SCHOOL*
FAM	1	0.959	0.986
PEER	–	1	0.983
SCHOOL	–	–	1

We see that the original OLS estimator for *SCHOOL* suggests a negative marginal effect, with respect to standard deviations, on students' achievements, holding other factors constant. However, our approach suggests a more reasonably positive marginal effect for *SCHOOL*, as summarized in Table 5. The confliction between the original and pertured results suggest that more investigation may be needed to confirm how those variables are related (Table 4).

Table 5 The summary of the original information for the education data, derived from the proposed algorithm

Variables	*FAM*	*PEER*	*SCHOOL*	*CN*	R^2
$\widehat{\beta}_{\mathcal{W}}$	0.46	0.54	0.14	14.39	0.18
*VIF*s	**2.66**	**2.85**	**3.84**	Iter	CPU
SE	3.33	3.44	3.99	87	1.73

Significance of bold shows the reduction in VIFs

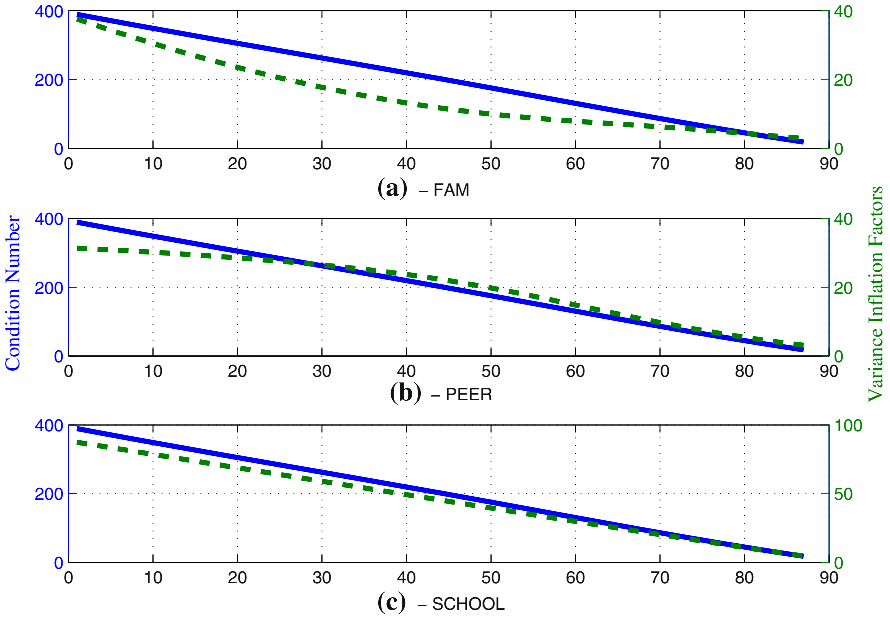


Fig. 3 The relationships between the condition number and *VIF*s for: **a** FAM, **b** PEER and **c** SCHOOL for education data. (*Solid line*—left axis; *dashed line*—right axis)

Figure 3 shows that all *VIF*s,together with the *CN*, decreases. More to that, if we compare (c) to (a) and (b) in Fig. 3, the proposed algorithm effectively renders the most pathological *VIF* to decrease stably over the iterations.

The next figure shows (i) the relationship between the condition number and the perturbation norm; and (ii) the relationship between the minimal and maximal eigenvalues for Ω^*_{XX} and $\mathcal{W}$. In Fig. 3a, the figure again confirms that, as the condition number drops down, the perturbation norm becomes larger. In Fig. 3b the positive definiteness is firm accordingly (Fig. 4).

6.3 The sales dataset

The following data represents a period of 23 years during which the firm was operating under fairly stable condition. The data shows the effect of the advertising expenditures

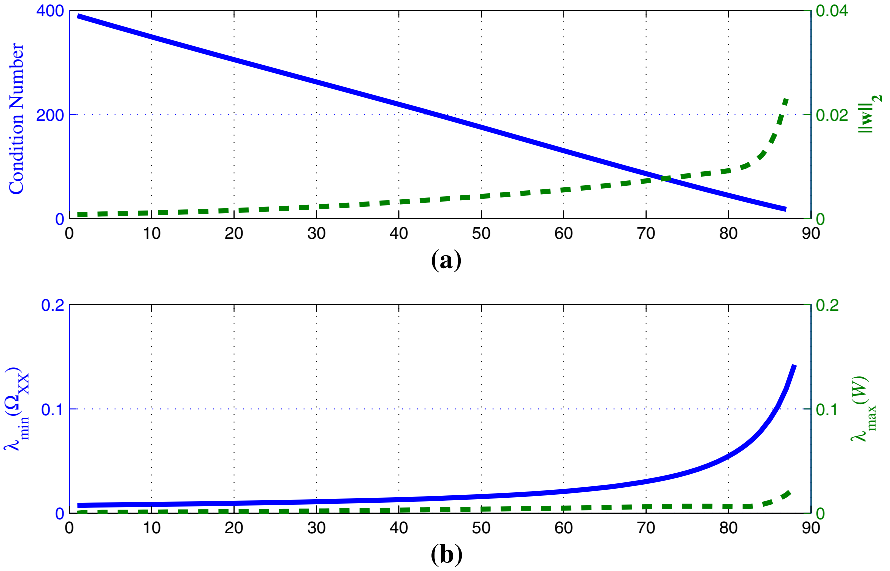


Fig. 4 **a** Relationship between condition number and the perturbation norm. **b** Concerns the positive definiteness of Ω^*_{XX} for education data. (*Solid line*—left axis; *dashed line*—right axis)

Table 6 The summary of the original information for the advertising data, derived from the ordinary least squares

Vars	A_t	P_t	E_t	A_{t-1}	P_{t-1}	CN	R^2
$\widehat{\beta}_{OLS}$	2.35	3.9	3.15	1.6	2.03	233.92	0.92
*VIF*s	**36.94**	**33.47**	**1.08**	**25.92**	**43.52**	–	–
SE	6.81	6.48	1.16	5.7	7.39	–	–

Significance of bold shows the reduction in VIFs

Table 7 The summary of the new information for the advertising data, derived from the proposed algorithm

Vars	A_t	P_t	E_t	A_{t-1}	P_{t-1}	CN	R^2
$\widehat{\beta}_{\mathcal{W}}$	1.98	3.49	3.44	1.21	1.6	11.78	0.92
*VIF*s	**2.51**	**2.85**	**1.16**	**1.97**	**2.36**	Iters	CPU
SE	1.03	1.77	2.61	0.72	0.79	45	0.69

Significance of bold shows the reduction in VIFs

(A_t and the lagged A_{t-1}), promotion expenditures (P_t and the lagged P_{t-1}), and sales expense (E_t) on the aggregate sales of a firm in period t. For details of the data, please see Chatterjee and Hadi (2006).

We summarize the OLS information is given in Table 6.

Applying the proposed approach to the dataset, the results are summarized in Table 7. The CPU time is 0.7 s for 47 iterations.

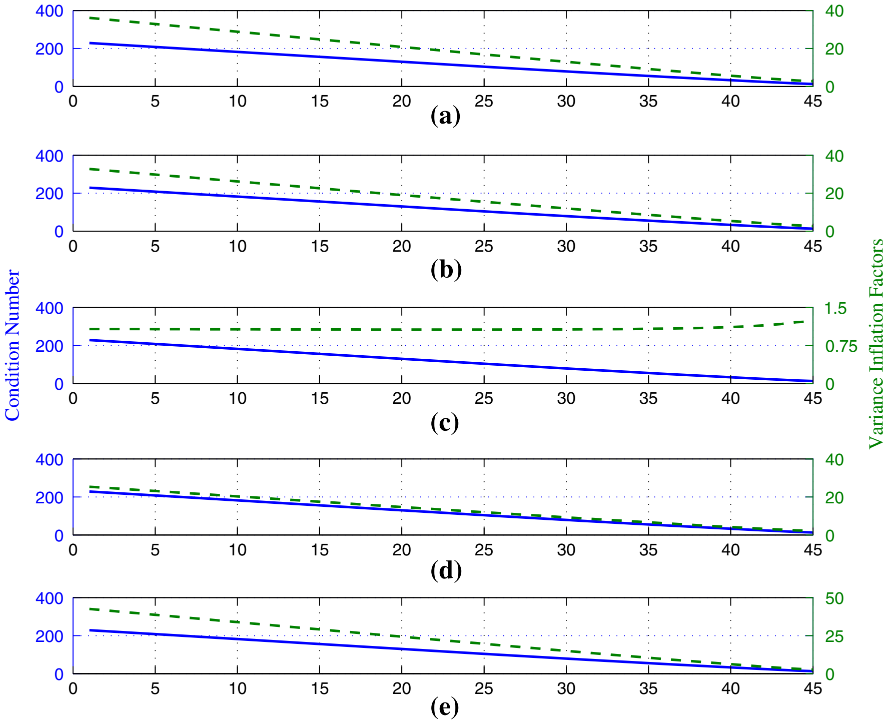


Fig. 5 The relationship between the condition number and *VIF*s for: **a** A_t, **b** P_t, **c** E_t, **d** A_{t-1} and **e** P_{t-1} (from top to bottom) (*solid line*—left axis; *dashed line*—right axis)

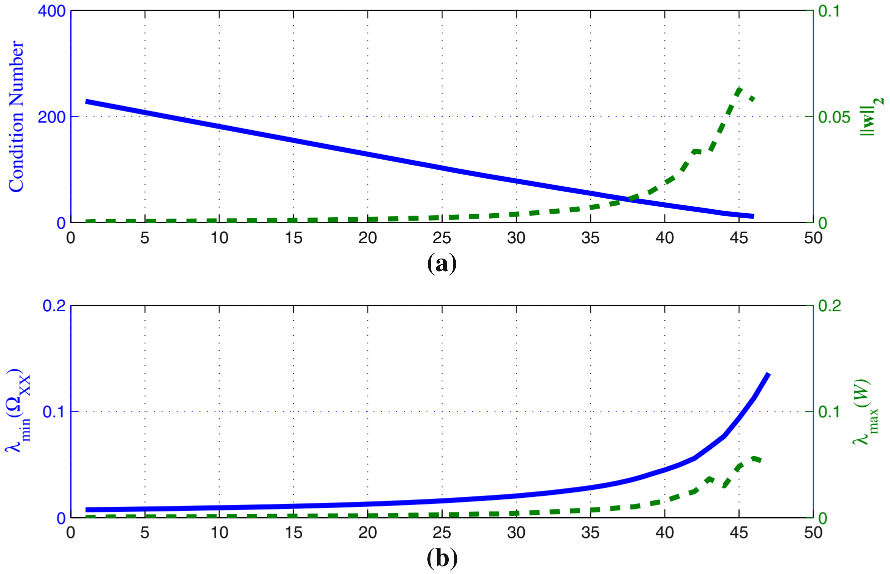


Fig. 6 **a** The relationship between the condition number and the perturbation norm; and **b** the positive definiteness of Ω^{*}_{XX} for sales data. (*Solid line*—left axis; *dashed line*—right axis)

Table 8 The correlation matrix for explanatory variables for US Pork Data

Vars	lp	lq	ldw	$ldpe$	$lcr4$	$d98$	t
lp	1	−0.25	0.89	−0.004	−0.79	−0.19	−0.86
lq	–	1	0.13	−0.17	−0.04	0.07	−0.11
ldw	–	–	1	−0.05	−0.78	−0.19	−0.91
$ldpe$	–	–	–	1	−0.02	−0.27	0.12
$lcr4$	–	–	–	–	1	0.25	0.95
$d98$	–	–	–	–	–	1	0.19
t	–	–	–	–	–	–	1

Table 9 The summary of the original information for (detrended) US pork data, derived from the ordinary least squares

Vars	lp	lq	ldw	$ldpe$	$lcr4$	$d98$	t	CN
$\widehat{\beta}_{OLS}$	0.19	−0.04	−0.15	0.06	0.06	−0.05	−0.43	336.74
*VIF*s	18.45	3.61	22.38	1.87	25.19	1.17	58.79	R^2
SE	0.28	0.13	0.31	0.09	0.33	0.07	0.51	0.98

Table 10 The summary of the original information for (trended) US pork data, derived from the ordinary least squares

Vars	lp	lq	ldw	$ldpe$	$lcr4$	$d98$
$\widehat{\beta}_{OLS}$	0.21	−0.03	−0.01	0.01	0.2	−0.05
*VIF*s	18.45	3.61	22.38	1.87	25.19	1.17
SE	0.37	0.16	0.33	0.09	0.16	0.09

In Fig. 5, we observe that the good *VIF* for (c) E_t fluctuates upward slightly, while other pathological *VIF*s drop down enormously. All results reveal the same conclusion as in the previous two examples (Fig. 6).

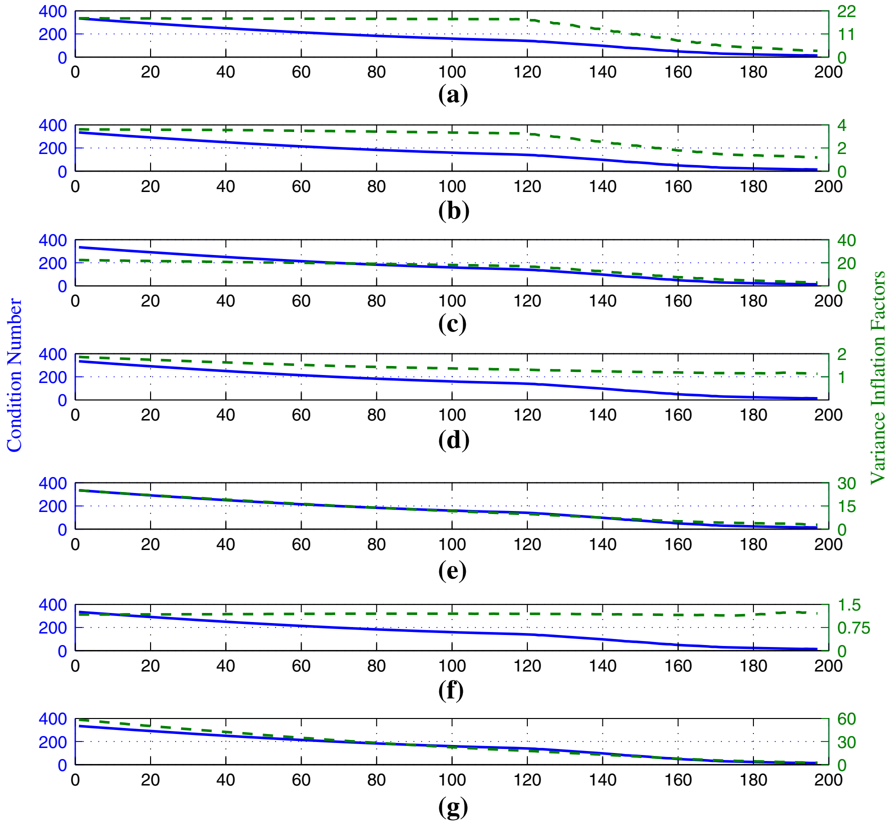
6.4 United States pork data

The following example illustrates the United States Pork dataset. The dataset contains the retail price (*rp*), deflated retail price (*drp*), net farm value (*nfv*), total price spread (*tps*), deflated price spread (*dps*), wage price (*wp*), wage price index (*wpi*), fuel price index (*fpi*), industrial cost (*ic*), deflated industrial cost (*dic*), quantity (*q*), consumer price index (*cpi*), and top 4 concentration rate (*cr4*) for US pork industry since 1970–2008 (Table 8).

We refer the detailed context of the dataset to Shen and Wohlgenant (2010). The model of interest is a detrended one, given by $lr = \beta_1 lp + \beta_2 lq + \beta_3 ldw + \beta_4 ldpe + \beta_5 lcr4 + \beta_6 d98 + \beta_7 t + \epsilon$, where lr is the log of the difference between deflated retail price and deflated price spread, lp is the log of deflated retail price, lq is the log of

Table 11 The summary of the original information for (detrended) US pork data, derived from the proposed algorithm

Vars	lp	lq	ldw	$ldpe$	$lcr4$	$d98$	t	CN	R^2
$\widehat{\beta}_W$	0.19	−0.04	0.03	0.01	−0.15	−0.05	−0.14	14.1	0.96
VIFs	2.65	1.05	2.69	1.73	2.95	1.12	2.31	Iters	CPU
SE	0.19	0.12	0.19	0.15	0.2	0.12	0.18	197	3.53



(a)

(b)

(c)

(d)

(e)

(f)

(g)

Fig. 7 The relationship between the condition number and the VIFs for: **a** lp, **b** lq, **c** ldw, **d** $ldpe$, **e** $lcr4$, **f** $d98$ and **g** t (from top to bottom). (*Solid line*—left axis; *dashed line*—right axis)

quantity, ldw is the log of deflated wage price index, $ldpe$ is the log of deflated fuel price index, $lcr4$ is the log os top four concentration rates, $d98$ is equal 1 if the year is 1998 or 1999; and 0, otherwise, and t is the cardinal of the years. As reported, the results for the trended version have a sign difference between coefficients for ldw. Too see that, we first look at the correlation matrix given below. Table 9 contains the OLS information. The trending variable was intended to account for the trending variations existing among the variables. However, it turns out that introducing a detrending variable t invokes serious collinearity problem for the problem.

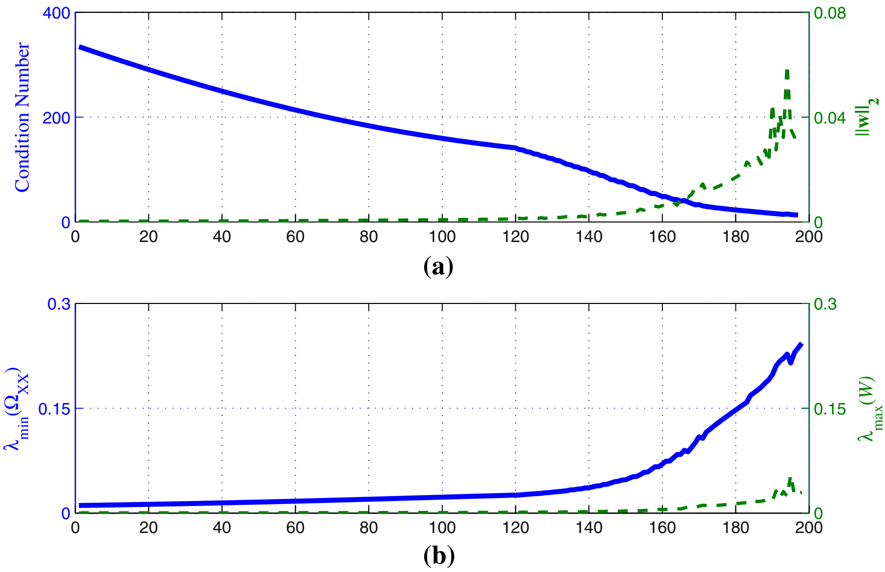


Fig. 8 **a** Relationship between the condition number and the perturbation norm; and **b** concerns the positive definiteness of Ω^*_{XX}. (*Solid line*—left axis; *dashed line*—right axis)

More to that, if we look at the trended version of the problem, we see that the coefficient for ldw has effectively a different sign, see Table 10.
We attempt to see if the proposed approach can mitigate the effect from introducing the detrending variable t. Table 11 suggests the perturbed information for the detrended model.

Two things can be observed. First, the proposed approach successfully reverses the sign for variable ldw, agreeing with that in the trended model. Second, however, as we see from Table 11, the coefficient for variable $lcr4$ becomes negative, meaning that the marginal percentage effect on lr has a 14 % drops, holding others constant.

Figure 7 reveals the fact that the *CN* behaves as *VIF*s, especially those pathological ones, do. In Fig. 8a, we observe that there are certain fluctuations on the curve for the perturbation norm over the iterations. Similarly, there is a quick incline in *CN* at around 120th iteration. The phenomena may be due to certain implicit numerical issues. By and large, Fig. 8a confirms that as the *CN* goes down, the perturbation norm goes up. Figure 8b simply confirms the positive definiteness of Ω_{XX}.

7 Concluding remarks

In this study, we propose a novel optimization model, based on the concept of *VIF*, to alleviate data collinearity problems in multiple linear regression. We show that the *VIF*s can decrease through solving the convex LCQP (2.15), using the proposed VIF-based algorithm. Various numerical examples validate the proposed approach.

The comparison between the proposed algorithm and the ridge regression can be unfair, because both approaches require exogenous parameters that directly affect the performance thereof. More to that, the relationship between the settings for ridge regression and our approach is not obviously related. In general, ridge regression outperforms the proposed algorithm in the *VIF* reduction.

There are a few issues needed to be solved. First of all, the development of the testing procedure for testing the hypothesis $H_0 : \left\{\varpi^*_{i,j} = \varpi_{i,j}\right\}$ against the alternative $H_1 : \left\{\varpi^*_{i,j} \neq \varpi_{i,j}\right\}$ is still missing in the literature. Once the testing procedure has been constructed, the examination of the perturbations generated by the LCQPs (2.15) can be done. Second, as the numerical examples suggest, the algorithm still suffer from numerical instability after certain iterations. This is due to the feature that the proposed algorithm can decrease the *CN* while decreasing the *VIF*s. As the *CN* has dropped down to a certain level, the sensitivity of the matrix becomes weak. So, a stablization of the performance of the algorithm may be needed.

Acknowledgments The authors thank the referees for their invaluable comments. The authors also thank those who have given their unconditional help in this paper.

References

Bagheri A, Midi H (2009) Robust estimations as a remedy for multicollinearity caused by multiple high leverage points. J Math Stat 5:311–321

Bashtian M, Arashi M, Tabatabaey SMM (2011) Using improved estimation strategies to combat multicollinearity. J Stat Comput Simul 81:1–25

Batah F, Özkale M, Gore S (2009) Combining unbiased ridge and principal component regression estimators. Commun Stat Theory 38:2201–2209

Bazaraa MS, Sherali HD, Shetty CM (2006) Nonlinear programming: theory and algorithms. Wiley, Hoboken

Belsley DA (1980) Regression diagnostics: identifying influential data and sources of collinearity. Wiley, Hoboken

Belsley DA (1984) Collinearity and forecasting. J Forecast 3:183–196

Chatterjee S, Hadi A (2006) Regression analysis by example. Wiley, Hoboken

Curto J, Pinto J (2007) New multicollinearity indicators in linear regression models. Int Stat Rev 75:114–121

Curto J, Pinto J (2011) The corrected VIF (CVIF). J Appl Stat 38:114–121

Echambadi R, Hess J (2007) Mean-centering does not alleviate collinearity problems in moderated multiple regression models. Market Sci 26:438–445

Efron B, Hastie T, Johnstone I, Tibshirani R (2004) Least angle regression. Ann Stat 32:407–499

Fierro R, Bunch J (1994) Collinearity and total least squares. SIAM J Matrix Anal Appl 15:1167–1181

Fox J, Monette G (1992) Generalized collinearity diagnostics. J Am Stat Assoc 87:178–183

Frank I, Friedman J (1993) A statistical view of some chemometrics regression tools (with discussion). Technometrics 35:109–148

Grewal R, Cote J, Baumgartner H (2004) Multicollinearity and measurement error in structural equation models: implications for theory testing. Market Sci 23:519–529

Hervé A (2010) Partial least squares regression and projection on latent structure regression (PLS Regression). Wiley Interdis Rev Comput Stat 2:97–106

Hoerl E, Kennard RW (1970) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12:55–67

Horn R, Johnson C (1990) Matrix analysis. Cambridge University Press, Cambridge

Knol DL, Ten Berge JMF (1989) Least-squares approximation of an improper correlation matrix by a proper one. Psychometrika 54:53–61

Kovás P, Petres T, Tóth L (2005) A new measure of multicollinearity in linear regression models. Int Stat Rev 73:405–412

Lazaridis A (2007) A note regarding the condition number: the case of spurious and latent multicollinearity. Qual Quant 41:123–135

Leung S, Yu S (2000) Collinearity and two-step estimation of sample selection models: problems, origins, and remedies. Comput Econ 15:173–199

Liao D, Valliant R (2012) Variance inflation factors in the analysis of complex survey data. Surv Method 38:53–62

Lin F (2008) Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. Qual Quant 42:417–426

Lin D, Foster D, Ungar L (2011) VIF regression: a fast regression algorithm for large data. J Am Stat Assoc 106:232–247

Lipovetsky S, Conklin W (2001) Multiobjective regression modifications for collinearity. Comput Oper Res 28:1333–1345

Liu K (2003) Using Liu-type estimator to combat collinearity. Commun Stat Theory 32:1009–1020

McDonald GC, Schwing RC (1973) Instabilities of regression estimates relating air pollution to mortality. Technom 15:463–481

McDonald GC (2009) Ridge regression. Wiley Interdis Rev Comput Stat 1:93–100

Miller KS (1981) On the inverse of the sum of matrices. Math Mag 54:67–72

Næs T, Mevik B (2001) Understanding the collinearity problem in regression and discriminant analysis. J Chemom 15:413–426

O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Qual Quant 41:673–690

Paul SR (1989) Test for the equality of several correlation coefficients. Can J Stat 17:217–227

Robinson C, Schumacher R (2009) Interaction effects: centering, variance inflation factor, and interpretation issues. Mult Linear Regres Viewp 35:6–11

Schwing RC, McDonald GC (1976) Measures of association of some air pollutants, natural ionizing radiation and cigarette smoking with mortality rates. Sci Total Environ 5:139–169

Shacham M, Brauner N (1997) Minimizing the effects of collinearity in polynomial regression. Ind Eng Chem Res 36:4405–4412

Shen Z, Wohlgenant M (2010) Modeling farm-retail price spread in the U.S. Pork Industry. Masters thesis, N.C. State University

Shieh G (2010) Clarifying the role of mean centring in multicollinearity of interaction effects. Brit J Math Stat Psychol. doi:10.1111/j.2044-8317.2010.02002.x

Soofi E (1990) Effects of collinearity on information about regression coefficients. J Econ 43:255–274

Spanos A, McGuirk A (2002) The problem of near-multicollinearity revisited: erractic vs systematic volatility. J Econ 108:365–393

Tibshirani R (1996) Regression shrinkage and selection via the Lasso. J R Stat Soc B 58:267–288

Xin L, Zhu M (2010) Stochastic stepwise ensembles for variable selection. J Comput Graph Stat 21:275–294