

## Research Article

# Revision: Variance Inflation in Regression

D. R. Jensen<sup>1</sup> and D. E. Ramirez<sup>2</sup>

<sup>1</sup> Department of Statistics, Virginia Tech, Blacksburg, VA 24061, USA

<sup>2</sup> Department of Mathematics, University of Virginia, P.O. Box 400137, Charlottesville, VA 22904, USA

Correspondence should be addressed to D. E. Ramirez; [der@virginia.edu](mailto:der@virginia.edu)

Received 12 October 2012; Accepted 10 December 2012

Academic Editor: Khosrow Moshirvaziri

Copyright © 2013 D. R. Jensen and D. E. Ramirez. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Variance Inflation Factors (VIFs) are reexamined as conditioning diagnostics for models with intercept, with and without centering regressors to their means as oft debated. Conventional VIFs, both centered and uncentered, are flawed. To rectify matters, two types of orthogonality are noted: vector-space orthogonality and uncorrelated centered regressors. The key to our approach lies in feasible Reference models encoding orthogonalities of these types. For models with intercept it is found that (i) uncentered VIFs are not ratios of variances as claimed, owing to infeasible Reference models; (ii) instead they supply informative angles between subspaces of regressors; (iii) centered VIFs are incomplete if not misleading, masking collinearity of regressors with the intercept; and (iv) *variance deflation* may occur, where ill-conditioned data yield smaller variances than their orthogonal surrogates. Conventional VIFs have all regressors linked, or none, often untenable in practice. Beyond these, our models enable the unlinking of regressors that can be unlinked, while preserving dependence among those intrinsically linked. Moreover, known collinearity indices are extended to encompass angles between subspaces of regressors. To reassess ill-conditioned data, we consider case studies ranging from elementary examples to data from the literature.

## 1. Introduction

Given  $\{Y = X\beta + \epsilon\}$  of full rank with zero-mean, uncorrelated and homoscedastic errors, the  $p$  equations  $\{X'X\beta = X'Y\}$  yield the OLS estimators  $\hat{\beta}$  for  $\beta = [\beta_1, \dots, \beta_p]'$  as unbiased with dispersion matrix  $V(\hat{\beta}) = \sigma^2 V$  and  $V = [v_{ij}] = (X'X)^{-1}$ . *Ill-conditioning*, as near dependent columns of  $X$ , exerts profound and interlinked consequences, causing “crucial elements of  $X'X$  to be large and unstable,” “creating inflated variances”; estimates excessive in magnitude, irregular in sign, and “very sensitive to small changes in  $X$ ”; and unstable algorithms having “degraded numerical accuracy.” See [1–3] for example.

Ill-conditioning diagnostics include the *condition number*  $c_1(X'X)$ , the ratio of its largest to smallest eigenvalues, and the *Variance Inflation Factors*  $\{VIF(\hat{\beta}_j) = u_{jj}/v_{jj}; 1 \leq j \leq p\}$  with  $U = X'X$ , that is, ratios of actual ( $v_{jj}$ ) to “ideal” ( $1/u_{jj}$ ) variances, had the columns of  $X$  been orthogonal. On scaling the latter to unit lengths and ordering  $\{VIF(\hat{\beta}_j) = v_{jj}; 1 \leq j \leq p\}$  as  $\{V_1 \geq V_2 \geq \dots \geq V_p\}$ ,  $V_1$  is identified in [4] as

“the best single measure of the conditioning of the data.” In addition, the bounds  $V_1 \leq c_1(X'X) \leq p(V_1 + \dots + V_p)$  of [5] apply also in stepwise regression as in [6–9].

Users deserve to be apprised not only that data are ill conditioned, but also about workings of the diagnostics themselves. Accordingly, we undertake here to rectify long-standing misconceptions in the use and properties of VIFs, necessarily retracing through several decades to their beginnings.

To choose a model, with or without intercept, is substantive, is specific to each experimental paradigm, and is beyond the scope of the present study. Whatever the choice, fixed in advance in a particular setting, these models follow on specializing  $p = k + 1$ , then  $p = k$ , respectively, as  $M_0 : \{Y = X_0\theta + \epsilon\}$  with intercept, and  $M_w : \{Y = X\beta + \epsilon\}$  without, where  $X_0 = [1_n, X]$ ;  $X = [X_1, X_2, \dots, X_k]$  comprise vectors of regressors; and  $\theta' = [\beta_0, \beta']$  with  $\beta_0$  as intercept and  $\beta' = [\beta_1, \dots, \beta_k]$  as slopes. The VIFs as defined are essentially undisputed for models in  $M_w$ , as noted in [10], serving to gage effects of nonorthogonal regressors

as ratios of variances. In contrast, a yet unresolved debate surrounds the choice of conditioning diagnostics for models in  $M_0$ , namely, between *uncentered* regressors giving  $VIF_u$ s for  $[\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k]$  and regressors  $\mathbf{X}_i$  *centered to their means*, giving  $VIF_c$ s for  $[\hat{\beta}_1, \dots, \hat{\beta}_k]$ . Specifically,  $\{VIF_u(\hat{\beta}_j) = u_{jj}v_{jj}; j = 0, 1, \dots, k\}$  on taking  $\mathbf{U} = \mathbf{X}'_0\mathbf{X}_0$  and  $\mathbf{V} = (\mathbf{X}'_0\mathbf{X}_0)^{-1}$ . In contrast, letting  $\mathbf{R} = [R_{ij}]$  be the correlation matrix for  $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1, \dots, \hat{\beta}_k]'$  and  $\mathbf{R}^{-1} = [R^{ij}]$  its inverse, then centered versions are  $\{VIF_c(\hat{\beta}_j) = R^{jj}; 1 \leq j \leq k\}$  for slopes only.

It is seen here that (i) these differ profoundly in regard to their respective concepts of orthogonality; (ii) objectives and meanings differ accordingly; (iii) sharp divisions trace to muddling these concepts; and (iv) this distinction assumes a pivotal role here. Over time a large body of widely held beliefs, conjectures, intrinsic propositions, and conventional wisdom has accumulated, much from flawed logic, some to be dispelled here. The key to our studies is that VIFs, to be meaningful, must compare actual variances to those of an “ideal” second-moment matrix as reference, the latter to embody the conjectured type of orthogonality. This differs between centered and uncentered diagnostics and for both types requires the reference matrix to be feasible. An outline follows.

Our undertaking consists essentially of four parts. The first is a literature survey of some essentials of ill-conditioning, to include the divide between centered and uncentered diagnostics and conventional VIFs. The structure of orthogonality is examined next. Anomalies in usage, meaning, and interpretation of conventional VIFs are exposed analytically and through elementary and transparent case studies. Long-standing but ostensible foundations in turn are reassessed and corrected through the construction of “Reference models.” These are moment arrays constrained to encode orthogonalities of the types considered. Neither array returns the conventional  $VIF_u$ s nor  $VIF_c$ s. Direct rules for finding the amended Reference models are given, preempting the need for constrained numerical algorithms. Finally, studies of ill-conditioned data from the literature are reexamined in light of these findings.

## 2. Preliminaries

**2.1. Notation.** Designate by  $\mathbb{R}^p$  the Euclidean  $p$ -space. Matrices and vectors are set in bold type; the transpose and inverse of  $\mathbf{A}$  are  $\mathbf{A}'$  and  $\mathbf{A}^{-1}$ ; and  $\mathbf{A}(ij)$  refers on occasion to the  $(i, j)$  element of  $\mathbf{A}$ . Special arrays are the identity  $\mathbf{I}_p$ ; the unit vector  $\mathbf{1}_p = [1, 1, \dots, 1]' \in \mathbb{R}^p$ ; the diagonal matrix  $\mathbf{D}_a = \text{Diag}(a_1, \dots, a_p)$ ; and  $\mathbf{B}_n = (\mathbf{I}_n - n^{-1}\mathbf{1}_n\mathbf{1}'_n)$ , as idempotent of rank  $n - 1$ . The Frobenius norm of  $\mathbf{x} \in \mathbb{R}^p$  is  $\|\mathbf{x}\| = (\mathbf{x}'\mathbf{x})^{1/2}$ . For  $\mathbf{X}$  of order  $(n \times p)$  and rank  $p \leq n$ , a generalized inverse is designated as  $\mathbf{X}^\dagger$ ; its ordered singular values as  $\sigma(\mathbf{X}) = \{\xi_1 \geq \xi_2 \geq \dots \geq \xi_p > 0\}$ ; and by  $\mathcal{S}_p(\mathbf{X}) \subset \mathbb{R}^n$  the subspace of  $\mathbb{R}^n$  spanned by the columns of  $\mathbf{X}$ . By accepted convention its condition number is  $c_2(\mathbf{X}) = \xi_1/\xi_p$ , specifically,  $c_2(\mathbf{X}) = [c_1(\mathbf{X}'\mathbf{X})]^{1/2}$ . For our model  $\{\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$ , with dispersion

$V(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}_n$ , we take  $\sigma^2 = 1.0$  unless stated otherwise, since variance ratios are scale invariant.

**2.2. Case Study 1: A First Look.** That anomalies pervade conventional  $VIF_u$ s and  $VIF_c$ s may be seen as follows. Given  $M_0 : \{Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_i\}$  the design  $\mathbf{X}_0 = [\mathbf{1}_5, \mathbf{X}_1, \mathbf{X}_2]$  of order  $(5 \times 3)$ ,  $\mathbf{U} = \mathbf{X}'_0\mathbf{X}_0$ , and its inverse  $\mathbf{V} = (\mathbf{X}'_0\mathbf{X}_0)^{-1}$  as in

$$\mathbf{X}'_0 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0.5 & 0.5 & 1 & 1 \\ -1 & 1 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} 5 & 3 & 1 \\ 3 & 2.5 & 1 \\ 1 & 1 & 3 \end{bmatrix}, \quad (1)$$

$$\mathbf{V} = \begin{bmatrix} 0.72222 & -0.88889 & 0.05556 \\ -0.88889 & 1.55556 & -0.22222 \\ 0.05556 & -0.22222 & 0.38889 \end{bmatrix}.$$

Conventional centered and uncentered VIFs are

$$[VIF_c(\hat{\beta}_1), VIF_c(\hat{\beta}_2)] = [1.0889, 1.0889], \quad (2)$$

$$[VIF_u(\hat{\beta}_0), VIF_u(\hat{\beta}_1), VIF_u(\hat{\beta}_2)] = [3.6111, 3.8889, 1.1667], \quad (3)$$

respectively, the former for slopes only and the latter taking reciprocals of the diagonals of  $\mathbf{X}'_0\mathbf{X}_0$  as reference.

**A Critique.** The following points are basic. Note first that model (1) is nonorthogonal in both the centered and uncentered regressors.

**Remark 1.** The  $VIF_u$ s are not ratios of variances and thus fail to gage relative increases in variances owing to nonorthogonal columns of  $\mathbf{X}_0$ . This follows since the first row and column of the second-moment matrix  $\mathbf{X}'_0\mathbf{X}_0 = \mathbf{U}$  are fixed and nonzero by design, so that taking  $\mathbf{X}'_0\mathbf{X}_0$  to be diagonal as reference cannot be feasible.

**Remark 1** runs contrary to assertions throughout the literature. In consequence, for models in  $M_0$  the mainstay  $VIF_u$ s in recent vogue are largely devoid of meaning. Subsequently these are identified instead with angles quantifying degrees of multicollinearity among the regressors.

On the other hand, feasible Reference models for all parameters, as identified later for centered and uncentered data in Definition 13, Section 4.2, give

$$[VF_c(\hat{\beta}_0), VF_c(\hat{\beta}_1), VF_c(\hat{\beta}_2)] = [0.9913, 1.0889, 1.0889], \quad (4)$$

$$[VF_v(\hat{\beta}_0), VF_v(\hat{\beta}_1), VF_v(\hat{\beta}_2)] = [0.7704, 0.8889, 0.8889] \quad (5)$$

in lieu of conventional  $VIF_c$ s and  $VIF_u$ s, respectively. The former comprise corrected  $VIF_c$ s, extended to include the intercept. Both sets in fact are genuine variance inflation factors, as ratios of variances in the model (1), relative to those in Reference models feasible for centered and for uncentered regressors, respectively.

This example flagrantly contravenes conventional wisdom: (i) variances for slopes are inflated in (4), but for the

intercept deflated, in comparison with the feasible centered reference. Specifically,  $\beta_0$  is estimated here with greater efficiency  $V\hat{F}_c(\beta_0) = 0.9913$  in the initial design (1), despite nonorthogonality of its centered regressors. (ii) Variances are uniformly smaller in the model (1) than in its feasible uncentered reference from (5), thus exhibiting *Variance Deflation*, despite nonorthogonality of the uncentered regressors. A full explication of the anatomy of this study appears later. In support, we next examine technical details needed in subsequent developments.

**2.3. Types of Orthogonality.** The ongoing divergence between centered and uncentered diagnostics traces in part to meaning ascribed to *orthogonality*. Specifically, the orthogonality of columns of  $\mathbf{X}$  in  $M_w$  refers unambiguously to the vector-space concept  $\mathbf{X}_i \perp \mathbf{X}_j$ , that is,  $\mathbf{X}'_i \mathbf{X}_j = 0$ , as does the notion of *collinearity of regressors with the constant vector* in  $M_0$ . We refer to this as *V-orthogonality*, in short  $V^\perp$ . In contrast, *nonorthogonality* in  $M_0$  often refers instead to the statistical concept of *correlation* among columns of  $\mathbf{X}$  when scaled and centered to their means. We refer to its negation as *C-orthogonality*, or  $C^\perp$ . Distinguishing between these notions is fundamental, as confusion otherwise is evident. For example, it is asserted in [11, p.125] that “the simple correlation coefficient  $r_{ij}$  does measure linear dependency between  $x_i$  and  $x_j$  in the data.”

**2.4. The Models  $M_0$ .** Consider  $M_0 : \{\mathbf{Y} = \mathbf{X}_0 \boldsymbol{\theta} + \boldsymbol{\epsilon}\}$ ,  $\mathbf{X}_0 = [\mathbf{1}_n, \mathbf{X}]$ ,  $\boldsymbol{\theta}' = [\beta_0, \boldsymbol{\beta}']$ , together with

$$\mathbf{X}'_0 \mathbf{X}_0 = \begin{bmatrix} n & n\bar{\mathbf{X}}' \\ n\bar{\mathbf{X}} & \mathbf{X}'\mathbf{X} \end{bmatrix}, \quad (\mathbf{X}'_0 \mathbf{X}_0)^{-1} = \begin{bmatrix} c_{00} & \mathbf{c}_{01} \\ \mathbf{c}'_{01} & \mathbf{G} \end{bmatrix}, \quad (6)$$

where  $n\bar{\mathbf{X}}' = \mathbf{1}'_n \mathbf{X}$ ,  $c_{00} = [n - n^2 \bar{\mathbf{X}}' (\mathbf{X}'\mathbf{X})^{-1} \bar{\mathbf{X}}]^{-1}$ ,  $\mathbf{C} = (\mathbf{X}'\mathbf{X} - n\bar{\mathbf{X}}\bar{\mathbf{X}}') = \mathbf{X}'\mathbf{B}_n\mathbf{X}$ , and  $\mathbf{G} = \mathbf{C}^{-1}$  from block-partitioned inverses. In later sections, we often will denote the mean-centering matrix  $n\bar{\mathbf{X}}\bar{\mathbf{X}}'$  by  $\mathbf{M}$ . In particular, the centered form  $\{\mathbf{X} \rightarrow \mathbf{Z} = \mathbf{B}_n\mathbf{X}\}$  arises exclusively in models with intercept, with or without reparametrizing  $(\beta_0, \boldsymbol{\beta}) \rightarrow (\alpha, \boldsymbol{\beta})$  in the mean-centered regressors, where  $\alpha = \beta_0 + \beta_1 \bar{X}_1 + \dots + \beta_k \bar{X}_k$ . Scaling  $\mathbf{Z}$  to unit column lengths gives  $\mathbf{Z}'\mathbf{Z}$  in correlation form with unit diagonals.

### 3. Historical Perspective

Our objective is an overdue revision of the tenets of variance inflation in regression. To provide context, we next survey extenuating issues from the literature. Direct quotations are intended not to subvert stances taken by the cited authors. Models in  $M_0$  are at issue since, as noted in [10], centered diagnostics have no place in  $M_w$ .

**3.1. Background.** Aspects of ill-conditioning span a considerable literature over many decades. Regarding  $\{\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$ , scaling columns of  $\mathbf{X}$  to equal lengths approximately minimizes the condition number  $c_2(\mathbf{X})$  [12, p.120] based on [13]. Nonetheless,  $c_2(\mathbf{X})$  is cast in [9] as a blunt instrument for

ill-conditioning, prompting the need for VIFs and other local constructs. Stewart [9] credits VIFs in concept to Daniel and in name to Marquardt.

Ill-conditioning points beyond OLS in view of difficulties cited earlier. Remedies proffered in [14, 15] include transforming variables, adding new data, and deleting variable(s) after checking critical features of the reduced model. Other intended palliatives include Ridge and Partial Least Squares, as compared in [16]; Principal Components regression; Surrogate models as in [17]. All are intended to return reduced standard errors at the expense of bias. Moreover, Surrogate solutions more closely resemble those from an orthogonal system than Ridge [18]. Together the foregoing and other options comprise a substantial literature as collateral to, but apart from, the present study.

**3.2. To Center.** Advocacy for centering includes the following.

- (i) VIFs often are *defined* categorically as the diagonals of the inverse of the correlation matrix of scaled and centered regressors; see [4, 9, 11, 19] for example. These are  $VIF_c$ s, widely adopted without justification as default to the exclusion of  $VIF_u$ s.
- (ii) It is asserted [4] that “centering removes the nonessential ill-conditioning, thus reducing the variance inflation in the coefficient estimates.”
- (iii) Centering is advocated when predictor variables are far removed from origins on the basic data scales [10, 11].

**3.3. Not to Center.** Advocacy for uncentered diagnostics includes the following caveats from proponents of centering.

- (i) Uncentered data should be examined only if an estimate of the intercept is of interest [9, 10, 20].
- (ii) “If the domain of prediction includes the full range from the natural origin through the range of data, the collinearity diagnostics should not be mean centered” [10, p.84].

Other issues against centering derive in part from numerical analysis and work by Belsley.

- (i) Belsley [1] identifies  $c_2(\mathbf{X})$  for a system  $\{\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$  as “the potential relative change in the LS solution  $\mathbf{b}$  that can result from a small relative change in the data.”
- (ii) These require *structurally interpretable* variables as “ones whose numerical values and (relative) differences derive meaning and interpretability from knowledge of the structure of the underlying ‘reality’ being modeled” [1, p.75].
- (iii) “There is no such thing as ‘nonessential’ ill-conditioning,” and “mean-centering can remove from the data the information needed to assess conditioning correctly” [1, p.74].
- (iv) “Collinearity with the intercept can quite generally corrupt the estimates of *all parameters* in the model

whether or not the intercept is itself of interest and whether or not the data have been centered” [21, p.90].

- (v) An example [22, p.121] gives severely ill-conditioned data perfectly conditioned in centered form: “centering alters neither” inflated variances nor extremely sensitive parameter estimates in the basic data; moreover, “diagnosing the conditioning of the centered data (which are perfectly conditioned) would completely overlook this situation, whereas diagnostics based on the basic data would not.”
- (vi) To continue from [22], ill-conditioning persists in the propagation of disturbances, in that “a 1 percent relative change in the  $\mathbf{X}_i$ ’s results in over a 40% relative change in the estimates,” despite perfect conditioning in centered form, and “knowledge of the effect of small relative changes in the centered data is not meaningful for assessing the sensitivity of the basic LS problem,” since relative changes and their meanings in centered and uncentered data often differ markedly.
- (vii) Regarding choice of origin, “(1) the investigator must be able to pick an origin against which small relative changes can be appropriately assessed and (2) it is the data measured relative to this origin that are relevant to diagnosing the conditioning of the LS problem” [22, p.126].

Other desiderata pertain.

- (i) “Because rewriting the model (in standardized variables) does not affect any of the implicit estimates, it has no effect on the amount of information contained in the data” [23, p.76].
- (ii) Consequences of ill-advised diagnostics can be severe. Degraded numerical accuracy traces to near collinearity of regressors with the constant vector. In short, centering fails to prevent a loss in numerical accuracy; centered diagnostics are unable to discern these potential accuracy problems, whereas uncentered diagnostics are seen to work well. Two widely used statistical packages, SAS and SPSS-X, fail to detect this type of ill-conditioning through use of centered diagnostics and thus return highly inaccurate coefficient estimates. For further details see [3].

On balance, for models in  $M_0$  the jury is out regarding the use of centered or uncentered diagnostics, to include VIFs. Even Belsley [1] (and elsewhere) concedes circumstances where centering does achieve structurally interpretable models. Of note is that the foregoing listed citations to Belsley apply strictly to condition numbers  $c_2(\mathbf{X})$ ; other purveyors of ill-conditioning, specifically VIFs, are not treated explicitly.

**3.4. A Synthesis.** It bears notice that (i) the origin, however remote from the cluster of regressors, is essential for prediction, and (ii) the prediction variance is invariant to parametrizing in centered or uncentered forms. Additional remarks are codified next for subsequent referral.

*Remark 2.* Typically  $Y$  represents response to input variables  $\{X_1, X_2, \dots, X_k\}$ . In a controlled experiment, levels are determined beforehand by subject-matter considerations extraneous to the experiment, to include minimal values. However remote the origin on the basic data scales, it seems informative in such circumstances to identify the origin with these minima. In such cases the intercept is often of singular interest, since  $\beta_0$  is then the standard against which changes in  $Y$  are to be gaged as regressors vary. We adopt this convention in subsequent studies from the archival literature.

*Remark 3.* In summary, the divergence in views, whether to center or not, appears to be that critical aspects of ill-conditioning, known and widely accepted for models in  $M_w$ , have been expropriated over decades, mindlessly and without verification, to apply point-by-point for models in  $M_0$ .

## 4. The Structure of Orthogonality

This section develops the foundations for Reference models capturing orthogonalities of types  $V^\perp$  and  $C^\perp$ . Essential collateral results are given in support as well.

**4.1. Collinearity Indices.** Stewart [9] reexamines ill-conditioning from the perspective of numerical analysis. Details follow, where  $\mathbf{X}$  is a generic matrix of regressors having columns  $\{\mathbf{x}_j; 1 \leq j \leq p\}$  and  $\mathbf{X}^\dagger = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  is the generalized inverse of note, having  $\{\mathbf{x}_j^\dagger; 1 \leq j \leq p\}$  as its typical rows. Each corresponding *collinearity index*  $\kappa$  is defined in [9, p.72] as

$$\{\kappa_j = \|\mathbf{x}_j\| \cdot \|\mathbf{x}_j^\dagger\|; 1 \leq j \leq p\}, \quad (7)$$

constructed so as to be scale invariant. Observe that  $\|\mathbf{x}_j^\dagger\|^2$  is found along the principal diagonal of  $[(\mathbf{X}^\dagger)(\mathbf{X}^\dagger)'] = (\mathbf{X}'\mathbf{X})^{-1}$ . When  $\mathbf{X}(n \times k)$  in  $\mathbf{X}_0 = [\mathbf{1}_n, \mathbf{X}]$  is centered and scaled, Section 3 of [9] shows that the *centered* collinearity indices  $\kappa_c$  satisfy  $\{\kappa_{c_j}^2 = \text{VIF}_c(\hat{\beta}_j); 1 \leq j \leq k\}$ . In  $M_0$  with parameters  $(\beta_0, \boldsymbol{\beta})$ , values corresponding to  $\|\mathbf{x}_j^\dagger\|^2$  from  $\mathbf{X}_0^\dagger = (\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'$  lie along the principal diagonal of  $(\mathbf{X}_0'\mathbf{X}_0)^{-1}$ ; the *uncentered* collinearity indices  $\kappa_u$  now satisfy  $\{\kappa_{u_j}^2 = \|\mathbf{x}_j\|^2 \cdot \|\mathbf{x}_j^\dagger\|^2; j = 0, 1, \dots, k\}$ . In particular, since  $\mathbf{x}_0 = \mathbf{1}_n$ , we have  $\kappa_{u_0}^2 = n\|\mathbf{x}_0^\dagger\|^2$ . Moreover, in  $M_0$  it follows that the uncentered  $\text{VIF}_u$ s are squares of the collinearity indices, that is,  $\{\text{VIF}_u(\hat{\beta}_j) = \kappa_{u_j}^2; j = 0, 1, \dots, k\}$ . Note the asymmetry that  $\text{VIF}_c$ s exclude the intercept, in contrast to the inclusive  $\text{VIF}_u$ s. That the label *Variance Inflation Factors* for the latter is a misnomer is covered in Remark 1. Nonetheless, we continue the familiar notation  $\{\text{VIF}_u(\hat{\beta}_j) = \kappa_{u_j}^2; j = 0, 1, \dots, k\}$ .

Transcending the essential developments of [9] are connections between collinearity indices and angles between

subspaces. To these ends choose a typical  $\mathbf{x}_j$  in  $\mathbf{X}_0$ , and rearrange  $\mathbf{X}_0$  as  $[\mathbf{x}_j, \mathbf{X}_{[j]}]$ . We next seek elements of

$$\mathbf{Q}'_j(\mathbf{X}'_0\mathbf{X}_0)^{-1}\mathbf{Q}_j = \begin{bmatrix} \mathbf{x}'_j\mathbf{x}_j & \mathbf{x}'_j\mathbf{X}_{[j]} \\ \mathbf{X}'_{[j]}\mathbf{x}_j & \mathbf{X}'_{[j]}\mathbf{X}_{[j]} \end{bmatrix}^{-1} \quad (8)$$

as reordered by the permutation matrix  $\mathbf{Q}_j$ . From the clockwise rule the (1, 1) element of the inverse is

$$[\mathbf{x}'_j(\mathbf{I}_n - \mathbf{P}_j)\mathbf{x}_j]^{-1} = [\mathbf{x}'_j\mathbf{x}_j - \mathbf{x}'_j\mathbf{P}_j\mathbf{x}_j]^{-1} = \|\mathbf{x}_j^\dagger\|^2 \quad (9)$$

in succession for each  $\{j = 0, 1, \dots, k\}$ , where  $\mathbf{P}_j = \mathbf{X}_{[j]}[\mathbf{X}'_{[j]}\mathbf{X}_{[j]}]^{-1}\mathbf{X}'_{[j]}$  is the projection operator onto the subspace  $\mathcal{S}_p(\mathbf{X}_{[j]}) \subset \mathbb{R}^n$  spanned by the columns of  $\mathbf{X}_{[j]}$ . These in turn enable us to connect  $\{\kappa_{uj}^2; j = 0, 1, \dots, k\}$ , and similarly for centered values  $\{\kappa_{cj}^2; j = 1, \dots, k\}$ , to the geometry of ill-conditioning as follows.

**Theorem 4.** For models in  $M_0$  let  $\{VIF_u(\hat{\beta}_j) = \kappa_{uj}^2; j = 0, 1, \dots, k\}$  be conventional uncentered  $VIF_u$ s in terms of Stewart's [9] uncentered collinearity indices. These in turn quantify the extent of collinearities between subspaces through angles (in deg) as follows.

- (i) Angles between  $(\mathbf{x}_j, \mathcal{S}_p(\mathbf{X}_{[j]}))$  are given by  $\theta_j = \arccos[(1 - 1/\kappa_{uj}^2)^{1/2}]$ , in succession for  $\{j = 0, 1, \dots, k\}$ .
- (ii) In particular,  $\theta_0 = \arccos[(1 - 1/\kappa_{u0}^2)^{1/2}]$  quantifies the degree of collinearity between the regressors and the constant vector.
- (iii) Similarly let  $\mathbf{B}_n\mathbf{X} = \mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_k]$  be regressors centered to their means, rearrange as  $[\mathbf{z}_j, \mathbf{Z}_{[j]}]$ , and let  $\{VIF_c(\hat{\beta}_j) = \kappa_{cj}^2; j = 1, \dots, k\}$  be centered  $VIF_c$ s in terms of Stewart's centered collinearity indices. Then angles (in deg) between  $(\mathbf{z}_j, \mathcal{S}_p(\mathbf{Z}_{[j]}))$  are given by  $\{\theta_j = \arccos[(1 - 1/\kappa_{cj}^2)^{1/2}]; 1 \leq j \leq k\}$ .

*Proof.* From the geometry of the right triangle formed by  $(\mathbf{x}_j, \mathbf{P}_j\mathbf{x}_j)$ , the squared lengths satisfy  $\|\mathbf{x}_j\|^2 = \|\mathbf{P}_j\mathbf{x}_j\|^2 + \|\mathbf{x}_j - \mathbf{P}_j\mathbf{x}_j\|^2$ , where  $\|\mathbf{x}_j - \mathbf{P}_j\mathbf{x}_j\|^2 = RS_j$  is the residual sum of squares from the projection. It follows that the principal angle between  $(\mathbf{x}_j, \mathbf{P}_j\mathbf{x}_j)$  is given by

$$\cos(\theta_j) = \frac{\mathbf{x}'_j\mathbf{P}_j\mathbf{x}_j}{\|\mathbf{x}_j\| \cdot \|\mathbf{P}_j\mathbf{x}_j\|} = \frac{\|\mathbf{P}_j\mathbf{x}_j\|}{\|\mathbf{x}_j\|} = \sqrt{1 - \frac{RS_j}{\|\mathbf{x}_j\|^2}} = \sqrt{1 - \frac{1}{\kappa_{uj}^2}} \quad (10)$$

for  $\{j = 0, 1, \dots, k\}$ , to give conclusion (i). Conclusion (ii) follows on specializing  $(\mathbf{x}_0, \mathbf{P}_0\mathbf{x}_0)$  with  $\mathbf{x}_0 = \mathbf{1}_n$  and  $\mathbf{P}_0 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ . Conclusion (iii) follows similarly from the geometry of the right triangle formed by  $(\mathbf{z}_j, \mathbf{P}_j\mathbf{z}_j)$ , where  $\mathbf{P}_j$  now is the projection operator onto the subspace  $\mathcal{S}_p(\mathbf{Z}_{[j]}) \subset \mathbb{R}^n$  spanned by the columns of  $\mathbf{Z}_{[j]}$ , to complete our proof.  $\square$

*Remark 5.* Rules of thumb in common use for problematic VIFs include those exceeding 10 or even 4; see [11, 24] for example. In angular measure these correspond, respectively, to  $\theta_j < 18.435$  deg and  $\theta_j < 30.0$  deg.

**4.2. Reference Models.** We seek as Reference feasible models encoding orthogonalities of types  $V^\perp$  and  $C^\perp$ . The keys are as follows: (i) to retain essentials of the experimental structure and (ii) to alter what may be changed to achieve orthogonality. For a model in  $M_w$  with moment matrix  $\mathbf{X}'\mathbf{X}$ , our opening paragraph prescribes as reference the model  $\mathbf{R}_V = \mathbf{D} = \text{Diag}(d_{11}, \dots, d_{pp})$ , as diagonal elements of  $\mathbf{X}'\mathbf{X}$ , for assessing  $V^\perp$ -orthogonality. Moreover, on scaling columns of  $\mathbf{X}$  to equal lengths,  $\mathbf{R}_V$  is perfectly conditioned with  $c_1(\mathbf{R}_V) = 1.0$ . In addition, every model in  $M_w$  clearly conforms with its Reference, in the sense that  $\mathbf{R}_V$  is positive definite, as distinct from models in  $M_0$  to follow.

Consider again models in  $M_0$  as in (6); let  $\mathbf{C} = (\mathbf{X}'\mathbf{X} - \mathbf{M})$  with  $\mathbf{M} = n\bar{\mathbf{X}}\bar{\mathbf{X}}'$  as the mean-centering matrix; and again let  $\mathbf{D} = \text{Diag}(d_{11}, \dots, d_{kk})$  comprise the diagonal elements of  $\mathbf{X}'\mathbf{X}$ .

(i)  $V^\perp$ -Reference Model. The uncentered  $VIF_u$ s in  $M_0$ , defined as ratios of diagonal elements of  $(\mathbf{X}'_0\mathbf{X}_0)^{-1}$  to reciprocals of diagonal elements of  $\mathbf{X}'_0\mathbf{X}_0$ , appear to have seen exclusive usage, apparently in keeping with Remark 3. However, the following disclaimer must be registered as the formal equivalent of Remark 1.

**Theorem 6.** Statements that conventional  $VIF_u$ s quantify variance inflation owing to nondiagonal  $\mathbf{X}'_0\mathbf{X}_0$  are false for models in  $M_0$  having  $\bar{\mathbf{X}} \neq \mathbf{0}$ .

*Proof.* Since the Reference variances are reciprocals of diagonal elements of  $\mathbf{X}'_0\mathbf{X}_0$ , this usage is predicated on the false postulate that  $\mathbf{X}'_0\mathbf{X}_0$  can be diagonal for  $\bar{\mathbf{X}} \neq \mathbf{0}$ . Specifically,  $[\mathbf{1}_n, \mathbf{X}]$  are linearly independent, that is,  $\mathbf{1}'_n\mathbf{X} = \mathbf{0}$ , if and only if  $\mathbf{X}$  has been mean centered beforehand.  $\square$

To the contrary, Gunst [25] purports to show that  $VIF_u(\hat{\beta}_0)$  registers genuine variance inflation, namely, the price to be paid in variance for designing an experiment having  $\bar{\mathbf{X}} \neq \mathbf{0}$ , as opposed to  $\bar{\mathbf{X}} = \mathbf{0}$ . Since variances for intercepts are  $\text{Var}(\cdot \mid \bar{\mathbf{X}} = \mathbf{0}) = \sigma^2/n$  and  $\text{Var}(\cdot \mid \bar{\mathbf{X}} \neq \mathbf{0}) = \sigma^2 c_{00} \geq \sigma^2/n$  from (6), their ratio is shown in [25] to be  $\kappa_{u0}^2 = nc_{00} \geq 1$  in the parlance of Section 2.3. We concede this to be a ratio of variances but, to the contrary, *not* a VIF, since the parameters differ. In particular,  $\text{Var}(\hat{\beta}_0 \mid \bar{\mathbf{X}} \neq \mathbf{0}) = \sigma^2 c_{00}$ , whereas  $\text{Var}(\hat{\alpha} \mid \bar{\mathbf{X}} \neq \mathbf{0}) = \sigma^2/n$ , with  $\alpha = \beta_0 + \beta_1\bar{X}_1 + \dots + \beta_k\bar{X}_k$  in centered regressors. Nonetheless, we still find the conventional  $VIF_u(\hat{\beta}_0)$  to be useful for purposes to follow.

*Remark 7.* Section 3 highlights continuing concern in regard to collinearity of regressors with the constant vector. Theorem 4(ii) and expression (10) support the use of  $\theta_0 = \arccos[(1 - 1/\kappa_{u0}^2)^{1/2}]$  as an informative gage on the extent of

this occurrence. Specifically, the smaller the angle, the greater the extent of such collinearity.

Instead of conventional VIF<sub>u</sub>s given the foregoing disclaimer, we have the following amended version as Reference for uncentered diagnostics, altering what may be changed but leaving  $\bar{\mathbf{X}} \neq \mathbf{0}$  intact.

**Definition 8.** Given a model in  $M_0$  with second-moment matrix  $\mathbf{X}'_0\mathbf{X}_0$ , the amended Reference model for assessing  $V^\perp$ -orthogonality is  $\mathbf{R}_V = \begin{bmatrix} n & n\bar{\mathbf{X}}' \\ n\bar{\mathbf{X}} & \mathbf{D} \end{bmatrix}$  with  $\mathbf{D} = \text{Diag}(d_{11}, \dots, d_{kk})$  as diagonal elements of  $\mathbf{X}'\mathbf{X}$ , provided that  $\mathbf{R}_V$  is positive definite. We identify a model to be  $V^\perp$ -orthogonal when  $\mathbf{X}'_0\mathbf{X}_0 = \mathbf{R}_V$ .

As anticipated, a prospective  $\mathbf{R}_V$  fails to conform to experimental data if not positive definite. These and further prospects are covered in the following, where  $\phi_i$  designates the angle between  $\{(\mathbf{1}_n, \mathbf{X}_i); i = 1, \dots, k\}$ .

**Lemma 9.** Take  $\mathbf{R}_V$  as a prospective Reference for  $V^\perp$ -orthogonality.

- (i) In order that  $\mathbf{R}_V$  maybe positive definite, it is necessary that  $\{n\bar{\mathbf{X}}'\mathbf{D}^{-1}\bar{\mathbf{X}} < 1\}$ , that is, that  $\{n \sum_{i=1}^k (\bar{X}_i^2/d_{ii}) < 1\}$ .
- (ii) Equivalently, it is necessary that  $\sum_{i=1}^k \cos^2(\phi_i) < 1$ , with  $\phi_i$  as the angle between  $\{(\mathbf{1}_n, \mathbf{X}_i); i = 1, \dots, k\}$ .
- (iii) The Reference variance for  $\hat{\beta}_0$  is  $\text{Var}(\hat{\beta}_0 | \mathbf{X}'\mathbf{X} = \mathbf{D}) = [n(1 - n\bar{\mathbf{X}}'\mathbf{D}^{-1}\bar{\mathbf{X}})]^{-1}$ .
- (iv) The Reference variances for slopes are given by

$$\left\{ \text{Var}(\hat{\beta}_i | \mathbf{X}'\mathbf{X} = \mathbf{D}) = \frac{1}{d_{ii}} \left[ 1 + \frac{\gamma \bar{X}_i^2}{d_{ii}} \right]; 1 \leq i \leq k \right\}, \quad (11)$$

where  $\gamma = n/(1 - n\bar{\mathbf{X}}'\mathbf{D}^{-1}\bar{\mathbf{X}})$ .

*Proof.* The clockwise rule for determinants gives  $|\mathbf{R}_V| = |\mathbf{D}||[n - n^2\bar{\mathbf{X}}'\mathbf{D}^{-1}\bar{\mathbf{X}}]|$ . Conclusion (i) follows since  $|\mathbf{D}| > 0$ . The computation

$$\cos(\phi_i) = \frac{\mathbf{1}'_n \mathbf{X}_i}{\|\mathbf{1}_n\| \|\mathbf{X}_i\|} = \frac{\mathbf{1}'_n \mathbf{X}_i}{\sqrt{nd_{ii}}} = \sqrt{\frac{n\bar{X}_i^2}{d_{ii}}}, \quad (12)$$

in parallel with (10), gives conclusion (ii). Using the clockwise rule for block-partitioned inverses, the (1, 1) element of  $\mathbf{R}_V^{-1}$  is given by conclusion (iii). Similarly, the lower right block of  $\mathbf{R}_V^{-1}$ , of order  $(k \times k)$ , is the inverse of  $\mathbf{H} = \mathbf{D} - n\bar{\mathbf{X}}\bar{\mathbf{X}}'$ . On identifying  $\mathbf{a} = \mathbf{b} = \bar{\mathbf{X}}$ ,  $\alpha = n$ , and  $\{a_i^* = a_i/d_{ii}\}$  in Theorem 8.3.3 of [26], we have that  $\mathbf{H}^{-1} = \mathbf{D}^{-1} + \gamma\mathbf{D}^{-1}\bar{\mathbf{X}}\bar{\mathbf{X}}'\mathbf{D}^{-1}$ . Conclusion (iv) follows on extracting its diagonal elements, to complete our proof.  $\square$

**Corollary 10.** For the case  $k = 2$ , in order that  $\mathbf{R}_V$  maybe positive definite, it is necessary that  $\phi_1 + \phi_2 > 90$  deg.

*Proof.* Beginning with Lemma 9(ii), compute

$$\begin{aligned} \cos(\phi_1 + \phi_2) &= \cos\phi_1 \cos\phi_2 - \sin\phi_1 \sin\phi_2 \\ &= \cos\phi_1 \cos\phi_2 - [1 - (\cos^2\phi_1 + \cos^2\phi_2) + \cos^2\phi_1 \cos^2\phi_2]^{1/2}, \end{aligned} \quad (13)$$

which is  $< 0$  when  $1 - (\cos^2\phi_1 + \cos^2\phi_2) > 0$ .  $\square$

Moreover, the matrix  $\mathbf{R}_V$  itself is intrinsically ill conditioned owing to  $\bar{\mathbf{X}} \neq \mathbf{0}$ , its condition number depending on  $\bar{\mathbf{X}}$ . To quantify this dependence, we have the following, where columns of  $\mathbf{X}$  have been standardized to common lengths  $\{|\mathbf{X}_i| = c; 1 \leq i \leq k\}$ .

**Lemma 11.** Let  $\mathbf{R}_V = \begin{bmatrix} n & \mathbf{T}' \\ \mathbf{T} & c\mathbf{I}_k \end{bmatrix}$  as in Definition 8, with  $\mathbf{T} = n\bar{\mathbf{X}}$ ,  $\mathbf{D} = c\mathbf{I}_k$ , and  $c > 0$ .

- (i) The ordered eigenvalues of  $\mathbf{R}_V$  are  $[\lambda_1, c, \dots, c, \lambda_p]$ , where  $p = k + 1$  and  $(\lambda_1, \lambda_p)$  are the roots of  $\lambda \in [(n+c) \pm \sqrt{(n-c)^2 + 4\tau^2}]/2$  and  $\tau^2 = \mathbf{T}'\mathbf{T}$ .
- (ii) The roots are positive, and  $\mathbf{R}_V$  is positive definite, if and only if  $(nc - \tau^2) > 0$ .
- (iii) If  $\mathbf{R}_V$  is positive definite, its condition number is  $c_1(\mathbf{R}_V) = \lambda_1/\lambda_p$  and is increasing in  $\tau^2$ .

*Proof.* Eigenvalues are roots of the determinantal equation

$$\begin{aligned} \text{Det} \begin{bmatrix} (n-\lambda) & \mathbf{T}' \\ \mathbf{T} & (c-\lambda)\mathbf{I}_k \end{bmatrix} &= 0 \iff (c-\lambda)^{k-1} [(n-\lambda)(c-\lambda) - \mathbf{T}'\mathbf{T}] = 0 \end{aligned} \quad (14)$$

from the clockwise rule, giving  $(k-1)$  values  $\{\lambda = c\}$  and two roots of the quadratic equation  $\{\lambda^2 - (n+c)\lambda + nc - \tau^2 = 0\}$ , to give conclusion (i). Conclusion (ii) holds since the product of roots of the quadratic equation is  $(nc - \tau^2)$  and the greater root is positive. Conclusion (iii) follows directly, to complete our proof.  $\square$

(ii)  $C^\perp$  Reference Model. As noted, the notion of  $C^\perp$ -orthogonality applies exclusively for models in  $M_0$ . Accordingly, as Reference we seek to alter  $\mathbf{X}'\mathbf{X}$  so that the matrix comprising sums of squares and sums of products of deviations from means, thus altered, is diagonal. To achieve this canon of  $C^\perp$ -orthogonality, and to anticipate notation for subsequent use, we have the following.

**Definition 12.** (i) For a model in  $M_0$  with second-moment matrix  $\mathbf{X}'_0\mathbf{X}_0$  and inverse as in (6), let  $\mathbf{C}_w = (\mathbf{W} - \mathbf{M})$ , and identify an indicator vector  $\mathbb{G} = [g_{12}, g_{13}, \dots, g_{1k}, g_{23}, \dots, g_{2k}, \dots, g_{k-1,k}]$  of order  $((k-1) \times (k-2))$  in lexicographic order, where  $\{g_{ij} = 0; i \neq j\}$  if the  $(i, j)$  element of  $\mathbf{C}_w$  is zero and  $\{g_{ij} = 1; i \neq j\}$  if the  $(i, j)$  element of  $\mathbf{C}_w$  is  $(\mathbf{X}'_i\mathbf{X}_j - n\bar{X}_i\bar{X}_j)$ , that is, unchanged from  $\mathbf{C} = (\mathbf{X}'\mathbf{X} - n\bar{\mathbf{X}}\bar{\mathbf{X}}') = \mathbf{Z}'\mathbf{Z}$ , with  $\mathbf{Z} = \mathbf{B}_n\mathbf{X}$  as the regressors centered to their means.

(ii) In particular, the Reference model in  $M_0$  for assessing  $C^\perp$ -orthogonality is  $\mathbf{R}_C = \begin{bmatrix} n & n\bar{\mathbf{x}}' \\ n\bar{\mathbf{x}} & \mathbf{W} \end{bmatrix}$  such that  $\mathbf{C}_w = (\mathbf{W} - \mathbf{M})$  and its inverse  $\mathbf{G}$  from (6) are diagonal, that is, taking  $\mathbf{W} = [w_{ij}]$  such that  $\{w_{ii} = d_{ii}; 1 \leq i \leq k\}$  and  $\{w_{ij} = m_{ij}; \text{ for all } i \neq j\}$  or equivalently,  $\mathbb{G} = [0, 0, \dots, 0]$ . In this case, we identify the model to be  $C^\perp$ -orthogonal.

Recall that conventional  $VIF_c$ s for  $[\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k]$  are ratios of diagonal elements of  $(\mathbf{Z}'\mathbf{Z})^{-1}$  to reciprocals of the diagonal elements of the centered  $\mathbf{Z}'\mathbf{Z}$ . Apparently this rests on the logic that  $C^\perp$ -orthogonality in  $M_0$  implies that  $\mathbf{Z}'\mathbf{Z}$  is diagonal. However, the converse fails and instead is embodied in  $\mathbf{R}_C \in M_0$ . In consequence, conventional  $VIF_c$ s are deficient in applying only to slopes, whereas the  $VIF_c$ s resulting from Definition 12(ii) apply informatively for all of  $[\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k]$ .

*Definition 13.* Designate  $VF_v(\cdot)$  and  $VF_c(\cdot)$  as *Variance Factors* resulting from Reference models of Definition 8 and Definition 12(ii), respectively. On occasion these represent *Variance Deflation* in addition to VIFs.

Essential invariance properties pertain. Conventional  $VIF_c$ s are scale invariant; see Section 3 of [9]. We see next that they are translation invariant as well. To these ends we shift the columns of  $\mathbf{X}=[\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k]$  to a new origin through  $\{\mathbf{X} \rightarrow \mathbf{X} - \mathbf{L} = \mathbf{H}\}$ , where  $\mathbf{L} = [c_1 \mathbf{1}_n, c_2 \mathbf{1}_n, \dots, c_k \mathbf{1}_n]$ . The resulting model is

$$\begin{aligned} \{\mathbf{Y} = (\beta_0 \mathbf{1}_n + \mathbf{L}\boldsymbol{\beta}) + \mathbf{H}\boldsymbol{\beta} + \boldsymbol{\epsilon}\} \\ \iff \{\mathbf{Y} = (\beta_0 + \psi) \mathbf{1}_n + \mathbf{H}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}, \end{aligned} \tag{15}$$

thus preserving slopes, where  $\psi = \mathbf{c}'\boldsymbol{\beta}=(c_1\beta_1+c_2\beta_2+\dots+c_k\beta_k)$ . Corresponding to  $\mathbf{X}_0 = [\mathbf{1}_n, \mathbf{X}]$  is  $\mathbf{U}_0 = [\mathbf{1}_n, \mathbf{H}]$  and to  $\mathbf{X}'_0\mathbf{X}_0$  and its inverse are

$$\mathbf{U}'_0\mathbf{U}_0 = \begin{bmatrix} n & \mathbf{1}'_n\mathbf{H} \\ \mathbf{H}'\mathbf{1}_n & \mathbf{H}'\mathbf{H} \end{bmatrix}, \quad (\mathbf{U}'_0\mathbf{U}_0)^{-1} = \begin{bmatrix} b_{00} & \mathbf{b}_{01} \\ \mathbf{b}'_{01} & \mathbf{G} \end{bmatrix} \tag{16}$$

in the form of (6). This pertains to subsequent developments, and basic invariance results emerge as follows.

**Lemma 14.** Consider  $\{\mathbf{Y} = \beta_0 \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$  together with the shifted version  $\{\mathbf{Y} = (\beta_0 + \psi) \mathbf{1}_n + \mathbf{H}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$ , both in  $M_0$ . Then the matrices  $\mathbf{G}$  appearing in (6) and (16) are identical.

*Proof.* Rules for block-partitioned inverses again assert that  $\mathbf{G}$  of (16) is the inverse of

$$\begin{aligned} (\mathbf{H}'\mathbf{H} - n^{-1}\mathbf{H}'\mathbf{1}_n\mathbf{1}'_n\mathbf{H}) &= \mathbf{H}'(\mathbf{I}_n - n^{-1}\mathbf{1}_n\mathbf{1}'_n)\mathbf{H} \\ &= \mathbf{H}'\mathbf{B}_n\mathbf{H} = \mathbf{X}'\mathbf{B}_n\mathbf{X} \end{aligned} \tag{17}$$

since  $\mathbf{B}_n\mathbf{L} = \mathbf{0}$ , to complete our proof.  $\square$

These facts in turn support subsequent claims that centered  $VIF_c$ s are translation and scale invariant for slopes  $\hat{\boldsymbol{\beta}}' = [\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k]$ , apart from the intercept  $\hat{\beta}_0$ .

4.3. A Critique. Again we distinguish  $VIF_c$ s and  $VIF_u$ s from centered and uncentered regressors. The following comments apply.

- (C1) A design is either  $V^\perp$  or  $C^\perp$ -orthogonal, respectively, according as the lower right  $(k \times k)$  block  $\mathbf{X}'\mathbf{X}$  of  $\mathbf{X}'_0\mathbf{X}_0$ , or  $\mathbf{C} = (\mathbf{X}'\mathbf{X} - n\bar{\mathbf{X}}\bar{\mathbf{X}}')$  from expression (6), is diagonal. Orthogonalities of type  $V^\perp$  and  $C^\perp$  are exclusive and hence work at crossed purposes.
- (C2) In particular,  $C^\perp$ -orthogonality holds if and only if the columns of  $\mathbf{X}$  are  $V^\perp$ -nonorthogonal. If  $\mathbf{X}$  is  $C^\perp$ -orthogonal, then for  $i \neq j$ ,  $\mathbf{C}_w(ij) = \mathbf{X}'_i\mathbf{X}_j - n\bar{X}_i\bar{X}_j = 0$  and  $\mathbf{X}'_i\mathbf{X}_j \neq 0$ . Conversely, if  $\mathbf{X}$  indeed is  $V^\perp$ -orthogonal, so that  $\mathbf{X}'\mathbf{X} = \mathbf{D} = \text{Diag}(d_{11}, \dots, d_{kk})$ , then  $(\mathbf{D} - n\bar{\mathbf{X}}\bar{\mathbf{X}}')$  cannot be diagonal as Reference, in which case the conventional  $VIF_c$ s are tenuous.
- (C3) Conventional  $VIF_u$ s, based on uncentered  $\mathbf{X}$  in  $M_0$  with  $\bar{\mathbf{X}} \neq \mathbf{0}$ , do not gage variance inflation as claimed, founded on the false tenet that  $\mathbf{X}'_0\mathbf{X}_0$  can be diagonal.
- (C4) To detect influential observations and to classify high leverage points, case-influence diagnostics, namely  $\{f_{jI} = VIF_{j(-I)}/VIF_j; 1 \leq j \leq k\}$ , are studied in [27] for assessing the impact of subsets on variance inflation. Here  $VIF_j$  is from the full data and  $VIF_{j(-I)}$  on deleting observations in the index set  $I$ . Similarly, [28] proposes using  $VIF_{(i)}$  on deleting the  $i$ th observation. In the present context these would gain substance on modifying  $VF_v$  and  $VF_c$  accordingly.

## 5. Case Study 1: Continued

5.1. The Setting. We continue an elementary and transparent example to illustrate essentials. Recall  $M_0 : \{Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_i\}$  of Section 2.2, the design  $\mathbf{X}_0 = [\mathbf{1}_5, \mathbf{X}_1, \mathbf{X}_2]$  of order  $(5 \times 3)$ , and  $\mathbf{U} = \mathbf{X}'_0\mathbf{X}_0$  and its inverse  $\mathbf{V} = (\mathbf{X}'_0\mathbf{X}_0)^{-1}$ , as in expressions (1). The design is neither  $V^\perp$  nor  $C^\perp$  orthogonal since neither the lower right  $(2 \times 2)$  block of  $\mathbf{X}'_0\mathbf{X}_0$  nor the centered  $\mathbf{Z}'\mathbf{Z}$  is diagonal. Moreover, the uncentered  $VIF_u$ s as listed in (3) are not the vaunted relative increases in variances owing to nonorthogonal columns of  $\mathbf{X}_0$ . Indeed, the only opportunity for  $V^\perp$ -orthogonality here is between columns  $[\mathbf{X}_1, \mathbf{X}_2]$ .

Nonetheless, from Section 4.1 we utilize Theorem 4(i) and (10) to connect the collinearity indices of [9] to angles between subspaces. Specifically, Minitab recovered values for the residuals  $\{RS_j = \|\mathbf{x}_j - \mathbf{P}_j\mathbf{x}_j\|^2; j = 0, 1, 2\}$ ; further computations proceed routinely as listed in Table 1. In particular, the principal angle  $\theta_0$  between the constant vector and the span of the regressor vectors is  $\theta_0 = 31.751$  deg, as anticipated in Remark 7.

5.2.  $V^\perp$ -Orthogonality. For subsequent reference let  $D(x)$  be a design as in (1) but with second-moment matrix:

$$\mathbf{U}(x) = \begin{bmatrix} 5 & 3 & 1 \\ 3 & 2.5 & x \\ 1 & x & 3 \end{bmatrix}. \tag{18}$$

TABLE 1: Values for residuals  $RS_j = \|\mathbf{x}_j - \mathbf{P}_j \mathbf{x}_j\|^2$ , for  $\|\mathbf{X}_j\|^2$ , for  $\kappa_{uj}^2 = \text{VIF}_u(\hat{\beta}_j)$ , for  $(1 - 1/\kappa_{uj}^2)^{1/2}$ , and for  $\theta_j$ , for  $j = 0, 1, 2$ .

$j$	$RS_j$	$\ \mathbf{X}_j\ ^2$	$\kappa_j^2$	$(1 - 1/\kappa_{uj}^2)^{1/2}$	$\theta_j$ (deg)
0	1.3846	5.0	3.6111	0.8503	31.751
1	0.6429	2.5	3.8889	0.8619	30.470
2	2.5714	3.0	1.1667	0.3780	67.792

Invoking Definition 8, we seek a  $V^\perp$ -orthogonal Reference with moment matrix  $\mathbf{U}(0)$  having  $\mathbf{X}'_1 \mathbf{X}_2 = 0$ , that is,  $D(0)$ . This is found constructively on retaining  $[\mathbf{1}_5, \bullet, \mathbf{X}_2]$  in  $\mathbf{X}_0$ , but replacing  $\mathbf{X}_1$  by  $\mathbf{X}'_1 = [1, 0.5, 0.5, 1, 0]'$ , as listed in Table 2, giving the design  $D(0)$  with columns  $[\mathbf{1}_5, \mathbf{X}'_1, \mathbf{X}_2]$  as Reference.

Accordingly,  $\mathbf{R}_V = \mathbf{U}(0)$  is “as  $V^\perp$ -orthogonal as it can get,” given the lengths and sums of  $(\mathbf{X}_1, \mathbf{X}_2)$  as prescribed in the experiment and as preserved in the Reference model. Lemma 9, with  $\bar{\mathbf{X}} = [0.6, 0.2]'$  and  $\mathbf{D} = \text{Diag}(2.5, 3)$ , gives  $\bar{\mathbf{X}}' \mathbf{D}^{-1} \bar{\mathbf{X}} = 0.157333$  and  $\gamma = 5/(1 - 5\bar{\mathbf{X}}' \mathbf{D}^{-1} \bar{\mathbf{X}}) = 23.4375$ . Applications of Lemma 9(iii)-(iv) in succession give the reference variances

$$\begin{aligned} \text{Var}(\hat{\beta}_0 | \mathbf{X}'\mathbf{X} = \mathbf{D}) &= \left[ 5 \left( 1 - 5\bar{\mathbf{X}}' \mathbf{D}^{-1} \bar{\mathbf{X}} \right) \right]^{-1} = 0.9375, \\ \text{Var}(\hat{\beta}_1 | \mathbf{X}'\mathbf{X} = \mathbf{D}) &= \frac{1}{d_{11}} \left[ 1 + \frac{\gamma \bar{X}_1^2}{d_{11}} \right] = \frac{2}{5} \left[ 1 + \frac{0.72\gamma}{5} \right] = 1.7500, \\ \text{Var}(\hat{\beta}_2 | \mathbf{X}'\mathbf{X} = \mathbf{D}) &= \frac{1}{d_{22}} \left[ 1 + \frac{\gamma \bar{X}_2^2}{d_{22}} \right] = \frac{1}{3} \left[ 1 + \frac{0.04\gamma}{3} \right] = 0.4375. \end{aligned} \quad (19)$$

As Reference, these combine with actual variances from  $\mathbf{V} = [V_{ij}]$  at (1), giving  $\text{VF}_v$  for the original design  $\mathbf{X}_0$  relative to  $\mathbf{R}_V$ . For example,  $\text{VF}_v(\hat{\beta}_0) = 0.7222/0.9375 = 0.7704$ , with  $V_{11} = 0.7222$  from (1), and 0.9375 from (19). This contrasts with  $\text{VIF}_u(\hat{\beta}_0) = 3.6111$  in (3) as in Section 2.2 and Table 2, it is noteworthy that the nonorthogonal design  $D(1)$  yields uniformly smaller variances than the  $V^\perp$ -orthogonal  $\mathbf{R}_V$ , namely,  $D(0)$ .

Further versions  $\{D(x); x \in [-0.5, 0.0, 0.5, 0.6, 1.0, 1.5]\}$  of our basic design are listed for comparison in Table 2, where  $x = \mathbf{X}'_1 \mathbf{X}_2$  for the various designs. The designs themselves are exhibited on transposing rows  $\mathbf{X}'_1 = [X_{11}, X_{12}, X_{13}, X_{14}, X_{15}]$  from Table 2, and substituting each into the template  $\mathbf{X}'_0 = [\mathbf{1}_5, \bullet, \mathbf{X}_2]$ . The design  $D(2)$  was constructed but not listed since its  $\mathbf{X}'_0 \mathbf{X}_0$  is not invertible. Clearly these are actual designs amenable to experimental deployment.

**5.3.  $C^\perp$ -Orthogonality.** Continuing  $\mathbf{X}_0 = [\mathbf{1}_5, \mathbf{X}_1, \mathbf{X}_2]$  and invoking Definition 12(ii), we seek a  $C^\perp$ -orthogonal Reference having the matrix  $\mathbf{C}_w$  as diagonal. This is

identified as  $D(0.6)$  in Table 2. From this the matrix  $\mathbf{R}_C$  and its inverse are

$$\begin{aligned} \mathbf{R}_C &= \begin{bmatrix} 5 & 3 & 1 \\ 3 & 2.5 & 0.6 \\ 1 & 0.6 & 3 \end{bmatrix}, \\ \mathbf{R}_C^{-1} &= \begin{bmatrix} 0.7286 & -0.8572 & -0.0714 \\ -0.8571 & 1.4286 & 0.0 \\ -0.0714 & 0.0 & 0.3571 \end{bmatrix}. \end{aligned} \quad (20)$$

The variance factors are listed in (4) where, for example,  $\text{VF}_c(\hat{\beta}_2) = 0.3889/0.3571 = 1.0889$ . As distinct from conventional  $\text{VIF}_c$ s for  $\hat{\beta}_1$  and  $\hat{\beta}_2$  only, our  $\text{VF}_c(\hat{\beta}_0)$  here reflects *Variance Deflation*, wherein  $\hat{\beta}_0$  is estimated more precisely in the initial  $C^\perp$ -nonorthogonal design.

A further note on orthogonality is germane. Suppose that the actual experiment is  $D(0)$  yet, towards a thorough diagnosis, the user evaluates the conventional  $\text{VIF}_c(\hat{\beta}_1) = 1.2250 = \text{VIF}_c(\hat{\beta}_2)$  as ratios of variances of  $D(0)$  to  $D(0.6)$  from Table 2. Unfortunately, their meaning is obscured since a design cannot at once be  $V^\perp$  and  $C^\perp$ -orthogonal.

As in Remark 2, we next alter the basic design  $\mathbf{X}_0$  at (1) on shifting columns of the measurements to  $[0, 0]$  as minima and scaling to have squared lengths equal to  $n = 5$ . The resulting  $\mathbf{X}_0$  follows directly from (1). The new matrix  $\mathbf{U} = \mathbf{X}'_0 \mathbf{X}_0$  and its inverse  $\mathbf{V}$  are

$$\begin{aligned} \mathbf{U} &= \begin{bmatrix} 5 & 4.2426 & 4.2426 \\ 4.2426 & 5 & 4 \\ 4.2426 & 4 & 5 \end{bmatrix}, \\ \mathbf{V} &= \begin{bmatrix} 1 & -0.4714 & -0.4714 \\ -0.4714 & 0.7778 & -0.2222 \\ -0.4714 & -0.2222 & 0.7778 \end{bmatrix} \end{aligned} \quad (21)$$

giving conventional  $\text{VIF}_u$ s as  $[5, 3.8889, 3.8889]$ . Against  $C^\perp$ -orthogonality, this gives the diagnostics

$$\left[ \text{VF}_c(\hat{\beta}_0), \text{VF}_c(\hat{\beta}_1), \text{VF}_c(\hat{\beta}_2) \right] = [0.8140, 1.0889, 1.0889] \quad (22)$$

demonstrating, in comparison with (4), that  $\text{VF}_c$ s, apart from  $\hat{\beta}_0$ , are invariant under translation and scaling, a consequence of Lemma 14.

We further seek to compare variances for the shifted and scaled  $(\mathbf{X}_1, \mathbf{X}_2)$  against  $V^\perp$ -orthogonality as Reference. However, from  $\mathbf{U} = \mathbf{X}'_0 \mathbf{X}_0$  at (21) we determine that  $\bar{X}_1 = 0.84853 = \bar{X}_2$  and  $5[(\bar{X}_1^2/5) + (\bar{X}_2^2/5)] = 1.4400$ . Since the latter exceeds unity, we infer from Lemma 9(i) that



TABLE 2: Designs  $\{D(x); x \in [-0.5, 0.0, 0.5, 0.6, 1.0, 1.5]\}$ , obtained by substituting  $\mathbf{X}_1^0 = [X_{11}, X_{12}, X_{13}, X_{14}, X_{15}]'$  in  $[\mathbf{I}_5, \bullet, \mathbf{X}_2]$  of order  $(5 \times 3)$ , and corresponding variances  $\{\text{Var}(\hat{\beta}_0), \text{Var}(\hat{\beta}_1), \text{Var}(\hat{\beta}_2)\}$ .

Design	$X_{11}$	$X_{12}$	$X_{13}$	$X_{14}$	$X_{15}$	$\text{Var}(\hat{\beta}_0)$	$\text{Var}(\hat{\beta}_1)$	$\text{Var}(\hat{\beta}_2)$
$D(-0.5)$	1.0	0.0	0.5	1.0	0.5	1.9333	3.7333	0.9333
$D(0.0)$	1.0	0.5	0.5	1.0	0.0	0.9375	1.7500	0.4375
$D(0.5)$	0.5	1.0	0.0	1.0	0.5	0.7436	1.4359	0.3590
$D(0.6)$	0.6091	0.5793	0.6298	1.1819	0.000	0.7286	1.4286	0.3570
$D(1.0)$	0.0	0.5	0.5	1.0	1.0	0.7222	1.5556	0.3889
$D(1.5)$	0.0	1.0	0.5	0.5	1.0	0.9130	2.4348	0.6087

$V^\perp$ -orthogonality is incompatible with this configuration of the data, so that the  $\text{VF}_v$ s are undefined. Equivalently, on evaluating  $\phi_1 = \phi_2 = 22.90$  deg and  $\phi_1 + \phi_2 = 45.80 < 90$  deg, Corollary 10 asserts that  $\mathbf{R}_V$  is not positive definite. This appears to be anomalous, until we recall that  $V^\perp$ -orthogonality is invariant under rescaling, but not recentering the regressors.

5.4. *A Critique.* We reiterate apparent ambiguity ascribed in the literature to *orthogonality*. A number of widely held guiding precepts has evolved, as enumerated in part in our opening paragraph and in Section 3. As in Remark 3, these clearly have been taken to apply verbatim for  $\mathbf{X}_0$  and  $\mathbf{X}_0' \mathbf{X}_0$  in  $M_0$ , to be paraphrased in part as follows.

- (P1) Ill-conditioning espouses inflated variances; that is, VIFs necessarily equal or exceed unity.
- (P2)  $C^\perp$ -orthogonal designs are “ideal” in that  $\text{VIF}_c$ s for such designs are all unity; see [11] for example.

We next reassess these precepts as they play out under  $V^\perp$  and  $C^\perp$  orthogonality in connection with Table 2.

- (C1) For models in  $M_0$  having  $\bar{\mathbf{X}} \neq \mathbf{0}$ , we reiterate that the uncentered  $\text{VIF}_u$ s at expression (3), namely  $[3.6111, 3.8889, 1.1667]$ , overstate adverse effects on variances of the  $V^\perp$ -nonorthogonal array, since  $\mathbf{X}_0' \mathbf{X}_0$  cannot be diagonal. Moreover, these conventional values fail to discern that the revised values for  $\text{VF}_v$ s at (5), namely,  $[0.7704, 0.8889, 0.8889]$ , reflect that  $[\beta_0, \beta_1, \beta_2]$  are estimated with *greater* efficiency in the  $V^\perp$ -nonorthogonal array  $D(1)$ .
- (C2) For  $V^\perp$ -orthogonality, the claim P1 thus is false by counterexample. The VFs at (5) are Variance Deflation Factors for design  $D(1)$ , where  $\mathbf{X}_1' \mathbf{X}_2 = 1$ , relative to the  $V^\perp$ -orthogonal  $D(0)$ .
- (C3) Additionally, the variances  $\{\text{Var}[\hat{\beta}_0(x)], \text{Var}[\hat{\beta}_1(x)], \text{Var}[\hat{\beta}_2(x)]\}$  in Table 2 are all seen to decrease from  $D(0) : [0.9375, 1.7500, 0.4375] \rightarrow D(1) : [0.7222, 1.5556, 0.3889]$  despite the transition from the  $V^\perp$ -orthogonal  $D(0)$  to the  $V^\perp$ -nonorthogonal  $D(1)$ . Similar trends are confirmed in other ill-conditioned data sets from the literature.
- (C4) For  $C^\perp$ -orthogonal designs, the claim P2 is false by counterexample. Such designs need not be “ideal,” as  $\beta_0$  may be estimated more efficiently in a  $C^\perp$ -nonorthogonal design, as demonstrated by the  $\text{VF}_c$ s

for  $\hat{\beta}_0$  at (4) and (22). Similar trends are confirmed elsewhere, as seen subsequently.

- (C5) VFs for  $\hat{\beta}_0$  are critical in prediction, where prediction variances necessarily depend on  $\text{Var}(\hat{\beta}_0)$ , especially for predicting near the origin of the system of coordinates.
- (C6) Dissonance between  $V^\perp$  and  $C^\perp$  is seen in Table 2, where  $D(0.6)$ , as  $C^\perp$ -orthogonal with  $\mathbf{X}_1' \mathbf{X}_2 = 0.6$ , is the antithesis of  $V^\perp$ -orthogonality at  $D(0)$ , where  $\mathbf{X}_1' \mathbf{X}_2 = 0$ .
- (C7) In short, these transparent examples serve to dispel the decades-old mantra that ill-conditioning necessarily spawns inflated variances for models in  $M_0$ , and they serve to illuminate the contributing structures.

## 6. Orthogonal and Linked Arrays

A genuine  $V^\perp$ -orthogonal array was generated as eigenvectors  $\mathbf{E} = [\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_8]$  from a positive definite  $(8 \times 8)$  matrix (see Table 8.3 of [11, p.377]) using PROC IML of the SAS programming package. The columns  $[\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$ , as the second through fourth columns of Table 3, comprise the first three eigenvectors scaled to length 8. These apply for models in  $M_w$  and  $M_0$  to be analyzed. In addition, linked vectors  $(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3)$  were constructed as  $\mathbf{Z}_1 = c(0.9\mathbf{E}_1 + 0.1\mathbf{E}_2)$ ,  $\mathbf{Z}_2 = c(0.9\mathbf{E}_2 + 0.1\mathbf{E}_3)$ , and  $\mathbf{Z}_3 = c(0.9\mathbf{E}_3 + 0.1\mathbf{E}_4)$ , where  $c = \sqrt{8/0.82}$ . Clearly these arrays, as listed in Table 3, are not archaic abstractions, as both are amenable to experimental implementation.

6.1. *The Model  $M_0$ .* We consider in turn the orthogonal and the linked series.

6.1.1. *Orthogonal Data.* Matrices  $\mathbf{X}_0' \mathbf{X}_0$  and  $(\mathbf{X}_0' \mathbf{X}_0)^{-1}$  for the orthogonal data under model  $M_0$  are listed in Table 4, where variances occupy diagonals of  $(\mathbf{X}_0' \mathbf{X}_0)^{-1}$ . The conventional uncentered  $\text{VIF}_u$ s are  $[1.20296, 1.02144, 1.12256, 1.05888]$ . Since  $\bar{\mathbf{X}} \neq \mathbf{0}$ , we find the angle between the constant and the span of the regressor vectors to be  $\theta_0 = 65.748$  deg as in Theorem 4(ii). Moreover, the angle between  $\mathbf{X}_1$  and the span of  $[\mathbf{I}_n, \mathbf{X}_2, \mathbf{X}_3]$ , namely,  $\theta_1 = 81.670$  deg, is not 90 deg because of collinearity with the constant vector, despite the mutual orthogonality of  $[\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$ . Observe here that  $\mathbf{X}_0' \mathbf{X}_0$

TABLE 3: Basic  $V^\perp$ -orthogonal design  $\mathbf{X}_0 = [\mathbf{1}_8, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$ , and a linked design  $\mathbf{Z}_0 = [\mathbf{1}_8, \mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3]$ , of order  $(8 \times 4)$ .

Orthogonal ( $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ )				Linked ( $\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3$ )			
$\mathbf{1}_8$	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{1}_8$	$\mathbf{Z}_1$	$\mathbf{Z}_2$	$\mathbf{Z}_3$
1	-1.084470	0.056899	0.541778	1	-1.071550	0.116381	0.653486
1	-1.090010	-0.040490	0.117026	1	-1.087820	-0.027320	0.116086
1	0.979050	0.002564	2.337454	1	0.973345	0.260677	2.199687
1	-1.090230	0.016839	0.170711	1	-1.081700	0.035588	0.070694
1	0.127352	2.821941	-0.071670	1	0.438204	2.796767	-0.070710
1	1.045409	-0.122670	-1.476870	1	1.025468	-0.285010	-1.618560
1	1.090803	-0.088970	0.043354	1	1.074306	-0.083640	0.176928
1	1.090739	-0.092300	0.108637	1	1.073875	-0.079740	0.244567

TABLE 4: Matrices  $\mathbf{X}'_0\mathbf{X}_0$  and  $(\mathbf{X}'_0\mathbf{X}_0)^{-1}$  for the orthogonal data under model  $M_0$ .

$\mathbf{X}'_0\mathbf{X}_0$				$(\mathbf{X}'_0\mathbf{X}_0)^{-1}$			
8	1.0686	2.5538	1.7704	0.1504	-0.0201	-0.0480	-0.0333
1.0686	8	0	0	-0.0201	0.1277	0.0064	0.0044
2.5538	0	8	0	-0.0480	0.0064	0.1403	0.0106
1.7704	0	0	8	-0.0333	0.0044	0.0106	0.1324

already is  $V^\perp$ -orthogonal; accordingly,  $\mathbf{R}_V = \mathbf{X}'_0\mathbf{X}_0$ ; and thus the  $\text{VF}_v$ s are all unity.

In view of dissonance between  $V^\perp$  and  $C^\perp$  orthogonality, the sums of squares and products for the mean-centered  $\mathbf{X}$  are

$$\mathbf{X}'\mathbf{B}_n\mathbf{X} = \begin{bmatrix} 7.8572 & -0.3411 & -0.2365 \\ -0.3411 & 7.1847 & -0.5652 \\ -0.2365 & -0.5652 & 7.6082 \end{bmatrix}, \quad (23)$$

reflecting nonnegligible negative dependencies, to distinguish the  $V^\perp$ -orthogonal  $\mathbf{X}$  from the  $C^\perp$ -nonorthogonal matrix  $\mathbf{Z} = \mathbf{B}_n\mathbf{X}$  of deviations.

To continue, we suppose instead that the  $V^\perp$ -orthogonal model was to be recast as  $C^\perp$ -orthogonal. The Reference model and its inverse are listed in Table 5. Variance factors, as ratios of diagonal elements of  $(\mathbf{X}'_0\mathbf{X}_0)^{-1}$  in Table 4 to those of  $\mathbf{R}_C^{-1}$  in Table 5, are

$$\begin{aligned} & [\text{VF}_c(\hat{\beta}_0), \text{VF}_c(\hat{\beta}_1), \text{VF}_c(\hat{\beta}_2), \text{VF}_c(\hat{\beta}_3)] \\ & = [1.0168, 1.0032, 1.0082, 1.0070], \end{aligned} \quad (24)$$

reflecting negligible differences in precision. This parallels results reported in Table 2 comparing  $D(0)$  to  $D(0.6)$  as Reference, except for larger differences in Table 2, namely,  $[1.2868, 1.2250, 1.2250]$ .

**6.1.2. Linked Data.** For the Linked Data under model  $M_0$ , the matrix  $\mathbf{Z}'_0\mathbf{Z}_0$  follows routinely from Table 3; its inverse  $(\mathbf{Z}'_0\mathbf{Z}_0)^{-1}$  is listed in Table 6. The conventional uncentered  $\text{VIF}_u$ s are now  $[1.20320, 1.03408, 1.13760, 1.05480]$ . These again fail to gage variance inflation since  $\mathbf{Z}'_0\mathbf{Z}_0$  cannot be diagonal. From (10) the principal angle between the constant vector and the span of the regressor vectors is  $\theta_0 = 65.735$  degrees.

*Remark 15.* Observe that the choice for  $\mathbf{R}_C$  may be posed as seeking  $\mathbf{R}_C = \mathbf{U}(x)$  such that  $\mathbf{R}_C^{-1}(2, 3) = 0.0$ , then solving numerically for  $x$  using Maple for example. However, the algorithm in Definition 12(ii) affords a direct solution: that  $\mathbf{C}_\psi$  should be diagonal stipulates its off-diagonal element as  $\mathbf{X}'_1\mathbf{X}_2 - 3/5 = 0$ , so that  $\mathbf{X}'_1\mathbf{X}_2 = 0.6$  in  $\mathbf{R}_C$  at expression (20).

To illustrate Theorem 4(iii), we compute  $\cos(\theta) = (1 - 1/1.0889)^{1/2} = 0.2857$  and  $\theta = 73.398$  deg as the angle between the vectors  $[\mathbf{X}_1, \mathbf{X}_2]$  of (1) when centered to their means. For slopes in  $M_0$ , [9] shows that  $\{\text{VIF}_c(\hat{\beta}_j) < \text{VIF}_u(\hat{\beta}_j); 1 \leq i \leq k\}$ . This follows since their numerators are equal, but denominators are reciprocals of lengths of the centered and uncentered regressor vectors. To illustrate, numerators are equal for  $\text{VIF}_u(\hat{\beta}_3) = 0.3889/(1/3) = 1.1667$  and for  $\text{VIF}_c(\hat{\beta}_3) = 0.3889/(1/2.8) = 1.0889$ , but denominators are reciprocals of  $\mathbf{X}'_2\mathbf{X}_2 = 3$  and  $\sum_{j=1}^5 (X_{2j} - \bar{X}_2)^2 = 2.8$ .

(i)  *$V^\perp$  Reference Model.* To rectify this malapropism, we seek in Definition 8 a model  $\mathbf{R}_V$  as Reference. This is found on setting all off-diagonal elements of  $\mathbf{Z}'_0\mathbf{Z}_0$  to zero, excluding the first row and column. Its inverse  $\mathbf{R}_V^{-1}$  is listed in Table 6. The  $\text{VF}_v$ s are ratios of diagonal elements of  $(\mathbf{Z}'_0\mathbf{Z}_0)^{-1}$  on the left to diagonal elements of  $\mathbf{R}_V^{-1}$  on the right, namely,

$$\begin{aligned} & [\text{VF}_v(\hat{\beta}_0), \text{VF}_v(\hat{\beta}_1), \text{VF}_v(\hat{\beta}_2), \text{VF}_v(\hat{\beta}_3)] \\ & = [0.9697, 0.9991, 0.9936, 0.9943]. \end{aligned} \quad (25)$$

Against conventional wisdom, it is counter-intuitive that  $[\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3]$  are all estimated with greater precision in the  $V^\perp$ -nonorthogonal model  $\mathbf{Z}'_0\mathbf{Z}_0$  than in the  $V^\perp$ -orthogonal model  $\mathbf{R}_V$  of Definition 8. As in Section 5.4, this serves again to refute the tenet that ill-conditioning espouses inflated

TABLE 5: Matrices  $\mathbf{R}_C$  and  $\mathbf{R}_C^{-1}$  for checking  $C^\perp$ -orthogonality in the orthogonal data  $\mathbf{X}_0$  under model  $M_0$ .

	$\mathbf{R}_C$			$\mathbf{R}_C^{-1}$			
8	1.0686	2.5538	1.7704	0.1479	-0.0170	-0.0444	-0.0291
1.0686	8	0.3411	0.2365	-0.0170	0.1273	0	0
2.5538	0.3411	8	0.5652	-0.0444	0	0.1392	0
1.7704	0.2365	0.5652	8	-0.0291	0	0	0.1314

TABLE 6: Matrices  $(\mathbf{Z}'_0\mathbf{Z}_0)^{-1}$  and  $\mathbf{R}_V^{-1}$  for checking  $V^\perp$ -orthogonality for the linked data under model  $M_0$ .

	$(\mathbf{Z}'_0\mathbf{Z}_0)^{-1}$			$\mathbf{R}_V^{-1}$			
0.1504	-0.0202	-0.0461	-0.0283	0.1551	-0.0261	-0.0530	-0.0344
-0.0202	0.1293	-0.00797	0.0053	-0.0261	0.1294	0.0089	0.0058
-0.0461	-0.0079	0.1422	-0.0054	-0.0530	0.0089	0.1431	0.0117
-0.0283	0.0053	-0.0054	0.1319	-0.0344	0.0058	0.0117	0.1326

variances for models in  $M_0$ , that is, that VFs necessarily equal or exceed unity.

(ii)  $C^\perp$  Reference Model. Further consider variances, were the Linked Data to be centered. As in Definition 12(ii), we seek a Reference model  $\mathbf{R}_C$  such that  $\mathbf{C}_w = (\mathbf{W} - \mathbf{M})$  is diagonal. The required  $\mathbf{R}_C$  and its inverse are listed in Table 7. Since variances in the Linked Data are [0.15040, 0.12926, 0.14220, 0.13185] from Table 6 and Reference values appear as diagonal elements of  $\mathbf{R}_C^{-1}$  on the right in Table 7, their ratios give the centered  $\text{VF}_c$ s as

$$\begin{aligned} & [\text{VF}_c(\hat{\beta}_0), \text{VF}_c(\hat{\beta}_1), \text{VF}_c(\hat{\beta}_2), \text{VF}_c(\hat{\beta}_3)] \\ & = [0.9920, 1.0049, 1.0047, 1.0030]. \end{aligned} \tag{26}$$

We infer that changes in precision would be negligible, if instead the Linked Data experiment was to be recast as a  $C^\perp$ -orthogonal experiment.

As in Remark 15, the reference matrix  $\mathbf{R}_C$  is constrained by the first and second moments from  $\mathbf{X}'_0\mathbf{X}_0$  and has free parameters  $\{x_1, x_2, x_3\}$  for the  $\{(2, 3), (2, 4), (3, 4)\}$  entries. The constraints for  $C^\perp$ -orthogonality are  $\{\mathbf{R}_C^{-1}(2, 3) = \mathbf{R}_C^{-1}(2, 4) = \mathbf{R}_C^{-1}(3, 4) = 0\}$ . Although this system of equations can be solved with numerical software such as Maple, the algorithm stated in Definition 12 easily yields the direct solution given here.

6.2. The Model  $M_w$ . For the Linked Data take  $\{Y_i = \beta_1 Z_{i1} + \beta_2 Z_{i2} + \beta_3 Z_{i3} + \epsilon_i; 1 \leq i \leq 8\}$  as  $\{\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$ , where the expected response at  $\{Z_1 = Z_2 = Z_3 = 0\}$  is  $E(Y) = 0.0$ , and there is no occasion to translate the regressors. The lower right  $(3 \times 3)$  submatrix of  $\mathbf{Z}'_0\mathbf{Z}_0(4 \times 4)$  from  $M_0$  is  $\mathbf{Z}'\mathbf{Z}$ ; its inverse is listed in Table 8. The ratios of diagonal elements of  $(\mathbf{Z}'\mathbf{Z})^{-1}$  to reciprocals of diagonal elements of  $\mathbf{Z}'\mathbf{Z}$  are the conventional uncentered  $\text{VIF}_u$ s, namely, [1.0123, 1.0247, 1.0123]. Equivalently, scaling  $\mathbf{Z}'\mathbf{Z}$  to have unit diagonals and inverting gives  $\mathbf{C}^{-1}$  as in Table 8; its diagonal elements are the  $\text{VIF}_u$ s. Throughout Section 6, these comprise the only correct interpretation of conventional  $\text{VIF}_u$ s as genuine variance ratios.

## 7. Case Study: Body Fat Data

7.1. The Setting. Body fat and morphogenic measurements were reported for  $n = 20$  human subjects in Table D.11 of [29, p.717]. Response and regressor variables are  $Y$ : amount of body fat;  $X_1$ : triceps skinfold thickness;  $X_2$ : thigh circumference; and  $X_3$ : mid-arm circumference, under  $M_0 : \{Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon_i\}$ . From the original data as reported the condition number of  $\mathbf{X}'_0\mathbf{X}_0$  is 1,039,931, and the  $\text{VIF}_u$ s are [271.4800, 446.9419, 63.0998, 77.8703]. On scaling columns of  $\mathbf{X}$  to equal column lengths  $\{\|\mathbf{X}_i\| = 5.0; i = 1, 2, 3\}$ , the resulting condition number of the revised  $\mathbf{X}'_0\mathbf{X}_0$  is 2,969.6518, and the uncentered  $\text{VIF}_u$ s are as before from invariance under scaling.

Against complaints that regressors are remote from their natural origins, we center regressors to  $[0, 0, 0]$  as origin on subtracting the minimal element from each column as in Remark 2, and scaling these to  $\{\|\mathbf{X}_i\| = 5.0; i = 1, 2, 3\}$ . This is motivated on grounds that centered diagnostics, apart from  $\text{VIF}_c(\hat{\beta}_0)$ , are invariant under shifts and scalings from Lemma 14. In addition, under the new origin  $[0, 0, 0]$ ,  $\beta_0$  assumes prominence as the baseline response against which changes due to regressors are to be gaged. The original data are available on the publisher's web page for [29].

Taking these shifted and scaled vectors as columns of  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$  in the revised  $\mathbf{X}_0 = [\mathbf{1}_n, \mathbf{X}]$ , values for  $\mathbf{X}'_0\mathbf{X}_0$  and its inverse are listed in Table 9. The condition number of  $\mathbf{X}'_0\mathbf{X}_0$  is now 113.6969 and its  $\text{VIF}_u$ s are [6.7756, 17.9987, 4.2782, 17.4484]. Additionally, from Theorem 4(ii) we recover the angle  $\theta_0 = 22.592$  deg to quantify collinearity of regressors with the constant vector.

In addition, the scaled and centered correlation matrix for  $\mathbf{X}$  is

$$\mathbf{R} = \begin{bmatrix} 1.0 & 0.0847 & 0.8781 \\ 0.0847 & 1.0 & 0.1424 \\ 0.8781 & 0.1424 & 1.0 \end{bmatrix} \tag{27}$$

indicating strong linkage between mean-centered versions of  $(\mathbf{X}_1, \mathbf{X}_3)$ . Moreover, the experimental design is neither  $V^\perp$  nor  $C^\perp$ -orthogonal, since neither the lower right  $(3 \times 3)$  block of  $\mathbf{X}'_0\mathbf{X}_0$  nor the centered matrix  $\mathbf{C} = (\mathbf{X}'\mathbf{X} - n\bar{\mathbf{X}}\bar{\mathbf{X}}')$  is diagonal.

TABLE 7: Matrices  $\mathbf{R}_C$  and  $\mathbf{R}_C^{-1}$  for checking  $C^\perp$ -orthogonality in the linked data under model  $M_0$ .

$\mathbf{R}_C$				$\mathbf{R}_C^{-1}$			
8	1.3441	2.7337	1.7722	0.1516	-0.0216	-0.0484	-0.0291
1.3441	8	0.4593	0.2977	-0.0216	0.1286	0	0
2.7337	0.4593	8	0.6056	-0.0484	0	0.1415	0
1.7722	0.2977	0.6056	8	-0.0291	0	0	0.1314

TABLE 8: Matrices  $(\mathbf{Z}'\mathbf{Z})^{-1}$  and  $\mathbf{C}^{-1}$  for the linked data under model  $M_w$ .

$(\mathbf{Z}'\mathbf{Z})^{-1}$			$\mathbf{C}^{-1}$		
0.1265	-0.0141	0.0015	1.0123	-0.1125	0.0123
-0.0141	0.1281	-0.0141	-0.1125	1.0247	-0.1125
0.0015	-0.0141	0.1265	0.0123	-0.1125	1.0123

(i)  $V^\perp$  Reference Model. To compare this with a  $V^\perp$ -orthogonal design, we turn again to Definition 8. However, evaluating the test criterion in Lemma 9(i) shows that this configuration is not commensurate with  $V^\perp$ -orthogonality, so that the  $\text{VF}_v$ s are undefined. Comparisons with  $C^\perp$ -orthogonal models are addressed next.

(ii)  $C^\perp$  Reference Model. As in Definition 12(ii), the centering matrix is

$$n\bar{\mathbf{X}}\bar{\mathbf{X}}' = \begin{bmatrix} 18.8889 & 18.9402 & 18.7499 \\ 18.9402 & 18.9916 & 18.8007 \\ 18.7499 & 18.8007 & 18.6118 \end{bmatrix}. \quad (28)$$

To check against  $C^\perp$ -orthogonality, we obtain the matrix  $\mathbf{W}$  of Definition 12(ii) on replacing off-diagonal elements of the lower right  $(3 \times 3)$  submatrix of  $\mathbf{X}'_0\mathbf{X}_0$  by corresponding off-diagonal elements of  $n\bar{\mathbf{X}}\bar{\mathbf{X}}'$ . The result is the matrix  $\mathbf{R}_C$  and its inverse as given in Table 10, where diagonal elements of  $\mathbf{R}_C^{-1}$  are the Reference variances. The  $\text{VF}_c$ s, found as ratios of diagonal elements of  $(\mathbf{X}'_0\mathbf{X}_0)^{-1}$  relative to those of  $\mathbf{R}_C^{-1}$ , are listed in Table 12 under  $\mathbb{G} = [0, 0, 0]$ , the indicator of Definition 12(i) for this case.

**7.2. Variance Factors and Linkage.** Traditional gages of ill-conditioning are patently absurd on occasion. By convention VIFs are “all or none,” wherein Reference models entail strictly diagonal components. In practice some regressors are inextricably linked: to get, or even visualize, orthogonal regressors may go beyond feasible experimental ranges, require extrapolation beyond practical limits, and challenge credulity. Response functions so constrained are described in [30] as “picket fences.” In such cases, taking  $\mathbf{C}$  to be diagonal as its  $C^\perp$  Reference is moot, at best an academic folly, abetted in turn by default in standard software packages. In short, conventional diagnostics here offer answers to irrelevant questions. Given pairs of regressors irrevocably linked, we seek instead to assess effects on variances for other pairs that could be unlinked by design. These comments apply especially to entries under  $\mathbb{G} = [0, 0, 0]$  in Table 12.

To proceed, we define *essential* ill-conditioning as regressors inextricably linked, necessarily to remain so and as *nonessential* if regressors could be unlinked by design. As a

followup to Section 3.2, for  $\bar{\mathbf{X}} \neq \mathbf{0}$  we infer that the constant vector is inextricably linked with regressors, thus accounting for essential ill-conditioning not removed by centering, contrary to claims in [4]. Indeed, this is the essence of Definition 8.

Fortunately, these limitations may be remedied through Reference models adapted to this purpose. This in turn exploits the special notation and trappings of Definition 12(i) as follows. In addition to  $\mathbb{G} = [0, 0, 0]$ , we iterate for indicators  $\mathbb{G} = [g_{12}, g_{13}, g_{23}]$  taking values in  $\{[1, 0, 0], [0, 1, 0], [0, 0, 1], [1, 1, 0], [1, 0, 1], [0, 1, 1]\}$ . Values of  $\text{VF}_c$ s thus obtained are listed in Table 12. To fix ideas, the Reference  $\mathbf{R}_C$  for  $\mathbb{G} = [1, 0, 0]$ , that is, for the constraint  $\mathbf{R}_C(2, 3) = \mathbf{X}'_0\mathbf{X}_0(2, 3) = 19.4533$ , is given in Table 11, together with its inverse. Found as ratios of diagonal elements of  $(\mathbf{X}'_0\mathbf{X}_0)^{-1}$  relative to those of  $\mathbf{R}_C^{-1}$  in Table 11,  $\text{VF}_c$ s are listed in Table 12 under  $\mathbb{G} = [1, 0, 0]$ . In short,  $\mathbf{R}_C$  is “as  $C^\perp$ -orthogonal as it could be” given the constraint. Other  $\text{VF}_c$ s in Table 12 proceed similarly. For all cases in Table 12, excluding  $\mathbb{G} = [0, 1, 1]$ ,  $\beta_0$  is estimated with greater efficiency in the original model than any of the postulated Reference models.

As in Remark 15, the reference matrix  $\mathbf{R}_C$  for  $\mathbb{G} = [1, 0, 0]$  is constrained by  $\mathbf{X}'_1\mathbf{X}_2 = 19.4533$ , to be retained, whereas  $\mathbf{X}'_1\mathbf{X}_3 = x_1$  and  $\mathbf{X}'_2\mathbf{X}_3 = x_2$  are to be determined from  $\{\mathbf{R}_C^{-1}(2, 4) = \mathbf{R}_C^{-1}(3, 4) = 0\}$ . Although this system can be solved numerically as noted, the algorithm stated in Definition 12 easily yields the direct solution given here.

Recall  $X_1$ : triceps skinfold thickness;  $X_2$ : thigh circumference; and  $X_3$ : mid-arm circumference; and from (27) that  $(X_1, X_3)$  are strongly linked. Invoking the cutoff rule [24] for VIFs in excess of 4.0, it is seen for  $\mathbb{G} = [0, 0, 0]$  in Table 12 that  $\text{VF}_c(\hat{\beta}_1)$  and  $\text{VF}_c(\hat{\beta}_3)$  are excessive, where  $[\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$  are forced to be mutually uncorrelated as Reference. Remarkably similar values are reported for  $\mathbb{G} \in \{[1, 0, 0], [0, 0, 1], [1, 0, 1]\}$ , all gaged against intractable Reference models wherein  $(X_1, X_3)$  are postulated to be uncorrelated, however, incredibly.

On the other hand,  $\text{VF}_c$ s for  $\mathbb{G} \in \{[0, 1, 0], [1, 1, 0], [0, 1, 1]\}$  in Table 12 are likewise comparable, where  $(\mathbf{X}_1, \mathbf{X}_3)$  are allowed to remain linked at  $\mathbf{X}'_1\mathbf{X}_3 = 24.2362$ . Here the  $\text{VF}_c$ s reflect negligible changes in efficiency of the estimates

TABLE 9: Matrices  $\mathbf{X}'_0\mathbf{X}_0$  and  $(\mathbf{X}'_0\mathbf{X}_0)^{-1}$  for the body fat data under model  $M_0$  centered to minima and scaled to equal lengths.

	$\mathbf{X}'_0\mathbf{X}_0$				$(\mathbf{X}'_0\mathbf{X}_0)^{-1}$			
20	19.4365	19.4893	19.2934	0.3388	-0.1284	-0.1482	-0.0203	
19.4365	25	19.4533	24.2362	-0.1284	0.7199	0.0299	-0.6224	
19.4893	19.4533	25	19.6832	-0.1482	0.0299	0.1711	-0.0494	
19.2934	24.2362	19.6832	25	-0.0203	-0.6224	-0.0494	0.6979	

TABLE 10: Matrices  $\mathbf{R}_C$  and  $\mathbf{R}_C^{-1}$  for the body fat data, adjusted to give off-diagonal elements the value 0.0, with code  $\mathbb{G} = [0, 0, 0]$ .

	$\mathbf{R}_C$				$\mathbf{R}_C^{-1}$			
20	19.4365	19.4893	19.2934	0.5083	-0.1590	-0.1622	-0.1510	
19.4365	25	18.9402	18.7499	-0.1590	0.1636	0	0	
19.4893	18.9402	25	18.8007	-0.1622	0	0.1664	0	
19.2934	18.7499	18.8007	25	-0.1510	0	0	0.1565	

in comparison with the original  $\mathbf{X}'_0\mathbf{X}_0$ , where  $[\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$  are all linked. In summary, negligible changes in overall efficiency would accrue on recasting the experiment so that  $(\mathbf{X}_1, \mathbf{X}_2)$  and  $(\mathbf{X}_2, \mathbf{X}_3)$  were pairwise uncorrelated.

Table 12 conveys further useful information on noting that VFs, as ratios, are multiplicative and divisible. For example, take the model coded  $\mathbb{G} = [1, 1, 0]$ , where  $\mathbf{X}'_1\mathbf{X}_2$  and  $\mathbf{X}'_1\mathbf{X}_3$  are retained at their initial values under  $\mathbf{X}'_0\mathbf{X}_0$ , namely, 19.4533 and 24.2362 from Table 9, with  $(\mathbf{X}_2, \mathbf{X}_3)$  unlinked. Now taking  $\mathbb{G} = [0, 1, 0]$  as reference, we extract the VFs on dividing elements in Table 12 at  $\mathbb{G} = [0, 1, 0]$ , by those of  $\mathbb{G} = [1, 1, 0]$ . The result is

$$\begin{aligned}
 & [\text{VF}(\hat{\beta}_0), \text{VF}(\hat{\beta}_1), \text{VF}(\hat{\beta}_2), \text{VF}(\hat{\beta}_3)] \\
 &= \left[ \frac{0.9196}{0.9848}, \frac{1.0073}{1.0057}, \frac{1.0282}{1.0208}, \frac{1.0208}{1.0208} \right] \quad (29) \\
 &= [0.9338, 1.0017, 1.0072, 1.0000].
 \end{aligned}$$

### 8. Conclusions

Our goal is clarity in the actual workings of VIFs as ill-conditioning diagnostics, thus to identify and rectify misconceptions regarding the use of VIFs for models  $\mathbf{X}_0 = [\mathbf{1}_n, \mathbf{X}]$  with intercept. Would-be arbiters for “correct” usage have divided between  $\text{VIF}_c$ s for mean-centered regressors and  $\text{VIF}_u$ s for uncentered data. We take issue with both. Conventional but clouded vision holds that (i)  $\text{VIF}_u$ s gage variance inflation owing to nonorthogonal columns of  $\mathbf{X}_0$  as vectors in  $\mathbb{R}^n$ ; (ii)  $\text{VIFs} \geq 1.0$ ; and (iii) models having orthogonal mean-centered regressors are “ideal” in possessing  $\text{VIF}_c$ s of value unity. Reprising Remark 3, these properties, widely and correctly understood for models in  $M_w$ , appear to have been expropriated without verification for models in  $M_0$  hence the anomalies.

Accordingly, we have distinguished vector-space orthogonality ( $V^\perp$ ) of columns of  $\mathbf{X}_0$ , in contrast to orthogonality ( $C^\perp$ ) of mean-centered columns of  $\mathbf{X}$ . A key feature is the construction of second-moment arrays as Reference models  $(\mathbf{R}_V, \mathbf{R}_C)$ , encoding orthogonalities of these types, and their Variance Factors as  $\text{VF}_v$ s and  $\text{VF}_c$ s, respectively. Contrary

to convention, we demonstrate analytically and through case studies that (i')  $\text{VIF}_u$ s do not gage variance inflation; (ii')  $\text{VFs} \geq 1.0$  need not hold, whereas  $\text{VF} < 1.0$  represents *Variance Deflation*; and (iii') models having orthogonal centered regressors are not necessarily “ideal.” In particular, variance deflation occurs when ill-conditioned data yield smaller variances than corresponding orthogonal surrogates.

In short, our studies call out to adjudicate the prolonged dispute regarding centered and uncentered diagnostics for models in  $M_0$ . Our summary conclusions follow.

- (S1) The choice of a model, with or without intercept, is a substantive matter in the context of each experimental paradigm. That choice is fixed beforehand in a particular setting, is structural, and is beyond the scope of the present study. Snee and Marquardt [10] correctly assert that VIFs are fully credited and remain undisputed in models *without* intercept: these quantities unambiguously retain the meanings originally intended, namely, as ratios of variances to gage effects of nonorthogonal regressors. For models *with* intercept the choice is between centered and uncentered diagnostics, which is to be weighed as follows.
- (S2) Conventional  $\text{VIF}_c$ s apply incompletely to slopes only, not necessarily grasping fully that a system is ill conditioned. Of particular concern is missing evidence on collinearity of regressors with the intercept. On the other hand, if  $C^\perp$ -orthogonality pertains, then  $\text{VF}_c$ s may supplant  $\text{VIF}_c$ s as applicable to all of  $[\beta_0, \beta_1, \dots, \beta_k]$ . Moreover, the latter may reveal that  $\text{VF}_c(\hat{\beta}_0) < 1.0$ , as in cited examples and of special import in predicting near the origin of the coordinate system.
- (S3) Our  $\text{VF}_v$ s capture correctly the concept apparently intended by conventional  $\text{VIF}_u$ s. Specifically, if  $V^\perp$ -orthogonality pertains, then  $\text{VF}_v$ s, as genuine ratios of variances, may supplant  $\text{VIF}_u$ s as now discredited gages for variance inflation.
- (S4) Returning to Section 3.2, we subscribe fully, but for different reasons, to Belsley’s and others’ contention

TABLE II: Matrices  $\mathbf{R}_C$  and  $\mathbf{R}_C^{-1}$  for the body fat data, adjusted to give off-diagonal elements the value 0.0 except  $\mathbf{X}'_1\mathbf{X}_2$ ; code  $\mathbb{G} = [1, 0, 0]$ .

$\mathbf{R}_C$				$\mathbf{R}_C^{-1}$			
20	19.4365	19.4893	19.2934	0.4839	-0.1465	-0.1497	-0.1510
19.4365	25	19.4533	18.7499	-0.1465	0.1648	-0.0141	0
19.4893	19.4533	25	18.8007	-0.1497	-0.0141	0.1676	0
19.2934	18.7499	18.8007	25	-0.1510	0	0	0.1565

TABLE 12: Summary  $\text{VF}_c$ s for the body fat data centered to minima and scaled to common lengths, identified with varying  $\mathbb{G} = [g_{12}, g_{13}, g_{23}]$  from Definition 12.

Elements of $\mathbb{G}$			$\text{VF}_c$ s				
$g_{12}$	$g_{13}$	$g_{23}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	
0	0	0	0.6665	4.3996	1.0282	4.4586	
1	0	0	0.7002	4.3681	1.0208	4.4586	
0	1	0	0.9196	1.0073	1.0282	1.0208	
0	0	1	0.7201	4.3996	1.0073	4.3681	
1	1	0	0.9848	1.0057	1.0208	1.0208	
1	0	1	0.7595	4.3681	1.0003	4.3681	
0	1	1	1.0248	1.0073	1.0073	1.0160	

that uncentered  $\text{VIF}_u$ s are essential in assessing ill-conditioned data. In the geometry of ill-conditioning, these should be retained in the pantheon of regression diagnostics, not as now debunked gages of variance inflation, but as more compelling angular measures of the collinearity of each column of  $\mathbf{X}_0 = [\mathbf{1}_n, \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k]$ , with the subspace spanned by remaining columns.

(S5) Specifically, the degree of collinearity of regressors with the intercept is quantified by  $\theta_0$  deg which, if small, is known to “corrupt the estimates of *all parameters* in the model whether or not the intercept is itself of interest and whether or not the data have been centered” [21, p.90]. This in turn would appear to elevate  $\theta_0$  in importance as a critical ill-conditioning diagnostic.

(S6) In contrast, Theorem 4(iii) identifies conventional  $\text{VIF}_c$ s with angles between a centered regressor and the subspace spanned by remaining centered regressors. For example,

$$\theta_1 = \arccos \left( \left[ 1 - \frac{1}{\text{VIF}_c(\hat{\beta}_1)} \right]^{1/2} \right) \quad (30)$$

is the angle between the centered vector  $\mathbf{Z}_1 = \mathbf{B}_n \mathbf{X}_1$  and the remaining centered vectors. These lie in the linear subspace of  $\mathbb{R}^n$  comprising the orthogonal complement to  $\mathbf{1}_n \in \mathbb{R}^n$  and thus bear on the geometry of ill-conditioning in this subspace.

(S7) Whereas conventional VIFs are “all or none,” to be gaged against strictly diagonal components, our Reference models enable unlinking what may be unlinked, while leaving other pairs of regressors linked. This enables us to assess effects on variances

attributed to allowable partial dependencies among some but not other regressors.

In conclusion, the foregoing summary points to limitations in modern software packages. *Minitab* v.15, *SPSS* v.19, and *SAS* v.9.2 are standard software packages which compute collinearity diagnostics returning, with the default options, the centered VIFs  $\{\text{VIF}_c(\hat{\beta}_j); 1 \leq j \leq k\}$ . To compute the uncentered  $\text{VIF}_u$ s, one needs to define a variable for the constant term and perform a linear regression using the options of fitting without an intercept. On the other hand, computations for  $\text{VF}_c$ s and  $\text{VF}_u$ s, as introduced here, require additional if straightforward programming, for example, Maple and PROC IML of the SAS System, as utilized here.

## References

- [1] D. A. Belsley, “Demeaning conditioning diagnostics through centering,” *The American Statistician*, vol. 38, no. 2, pp. 73–77, 1984.
- [2] E. R. Mansfield and B. P. Helms, “Detecting multicollinearity,” *The American Statistician*, vol. 36, no. 3, pp. 158–160, 1982.
- [3] S. D. Simon and J. P. Lesage, “The impact of collinearity involving the intercept term on the numerical accuracy of regression,” *Computer Science in Economics and Management*, vol. 1, no. 2, pp. 137–152, 1988.
- [4] D. W. Marquardt and R. D. Snee, “Ridge regression in practice,” *The American Statistician*, vol. 29, no. 1, pp. 3–20, 1975.
- [5] K. N. Berk, “Tolerance and condition in regression computations,” *Journal of the American Statistical Association*, vol. 72, no. 360, part 1, pp. 863–866, 1977.
- [6] A. E. Beaton, D. Rubin, and J. Barone, “The acceptability of regression solutions: another look at computational accuracy,” *Journal of the American Statistical Association*, vol. 71, no. 353, pp. 158–168, 1976.

- [7] R. B. Davies and B. Hutton, "The effect of errors in the independent variables in linear regression," *Biometrika*, vol. 62, no. 2, pp. 383–391, 1975.
- [8] D. W. Marquardt, "Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation," *Technometrics*, vol. 12, no. 3, pp. 591–612, 1970.
- [9] G. W. Stewart, "Collinearity and least squares regression," *Statistical Science*, vol. 2, no. 1, pp. 68–84, 1987.
- [10] R. D. Snee and D. W. Marquardt, "Comment: collinearity diagnostics depend on the domain of prediction, the model, and the data," *The American Statistician*, vol. 38, no. 2, pp. 83–87, 1984.
- [11] R. H. Myers, *Classical and Modern Regression with Applications*, PWS-Kent, Boston, Mass, USA, 2nd edition, 1990.
- [12] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, New York, NY, USA, 1980.
- [13] A. van der Sluis, "Condition numbers and equilibration of matrices," *Numerische Mathematik*, vol. 14, pp. 14–23, 1969.
- [14] G. C. S. Wang, "How to handle multicollinearity in regression modeling," *Journal of Business Forecasting Methods & System*, vol. 15, no. 1, pp. 23–27, 1996.
- [15] G. C. S. Wang and C. L. Jain, *Regression Analysis Modeling & Forecasting*, Graceway, Flushing, NY, USA, 2003.
- [16] N. Adnan, M. H. Ahmad, and R. Adnan, "A comparative study on some methods for handling multicollinearity problems," *Matematika UTM*, vol. 22, pp. 109–119, 2006.
- [17] D. R. Jensen and D. E. Ramirez, "Anomalies in the foundations of ridge regression," *International Statistical Review*, vol. 76, pp. 89–105, 2008.
- [18] D. R. Jensen and D. E. Ramirez, "Surrogate models in ill-conditioned systems," *Journal of Statistical Planning and Inference*, vol. 140, no. 7, pp. 2069–2077, 2010.
- [19] P. F. Velleman and R. E. Welsch, "Efficient computing of regression diagnostics," *The American Statistician*, vol. 35, no. 4, pp. 234–242, 1981.
- [20] R. F. Gunst, "Regression analysis with multicollinear predictor variables: definition, detection, and effects," *Communications in Statistics A*, vol. 12, no. 19, pp. 2217–2260, 1983.
- [21] G. W. Stewart, "Comment: well-conditioned collinearity indices," *Statistical Science*, vol. 2, no. 1, pp. 86–91, 1987.
- [22] D. A. Belsley, "Centering, the constant, first-differencing, and assessing conditioning," in *Model Reliability*, D. A. Belsley and E. Kuh, Eds., pp. 117–153, MIT Press, Cambridge, Mass, USA, 1986.
- [23] G. Smith and F. Campbell, "A critique of some ridge regression methods," *Journal of the American Statistical Association*, vol. 75, no. 369, pp. 74–103, 1980.
- [24] R. M. O'Brien, "A caution regarding rules of thumb for variance inflation factors," *Quality & Quantity*, vol. 41, no. 5, pp. 673–690, 2007.
- [25] R. F. Gunst, "Comment: toward a balanced assessment of collinearity diagnostics," *The American Statistician*, vol. 38, pp. 79–82, 1984.
- [26] F. A. Graybill, *Introduction to Matrices with Applications in Statistics*, Wadsworth, Belmont, Calif, USA, 1969.
- [27] D. Sengupta and P. Bhimasankaram, "On the roles of observations in collinearity in the linear model," *Journal of the American Statistical Association*, vol. 92, no. 439, pp. 1024–1032, 1997.
- [28] J. O. van Vuuren and W. J. Conradie, "Diagnostics and plots for identifying collinearity-influential observations and for classifying multivariate outliers," *South African Statistical Journal*, vol. 34, no. 2, pp. 135–161, 2000.
- [29] R. R. Hocking, *Methods and Applications of Linear Models*, John Wiley & Sons, Hoboken, NJ, USA, 2nd edition, 2003.
- [30] R. R. Hocking and O. J. Pendleton, "The regression dilemma," *Communications in Statistics*, vol. 12, pp. 497–527, 1983.