24th International Congress on Modelling and Simulation, Sydney, NSW, Australia, 5 to 10 December 2021
mssanz.org.au/modsim2021

Heteroscedasticity and correlation in linear regression

A. Gill [a]

[a] *Joint and Operations Analysis Division, Defence Science and Technology Group, Australia*
Email: andrew.gill@dst.defence.gov.au

Abstract: At the last MODSIM conference I illustrated two common pitfalls that may present themselves during the design and analysis of simulation experiments (Gill 2019). The first provided good reason to seek and employ orthogonal designs, such as two-level fractional factorials or orthogonal Latin Hypercubes (Montgomery (2012) provides a good introduction). The second pitfall belies the analysis of simulation experiments and the potential dangers of making the common *independent and identically distributed* (iid) assumptions on the regression residuals. Such assumptions allow the classic *Analysis of Variance* (ANOVA) statistical procedures to be employed, as taught in statistical texts such as Montgomery (2012) and others, and are often standard on statistical software (e.g., Minitab and JMP). However, in simulation experiments the assumption of *identically distributed* responses at each design point is often not met in practice. In fact, heteroscedasticity is more often the norm and Law (2007) provides examples where the variances can differ by an order of magnitude or more, and while our control over the assignment of the *pseudo-random number* (PRN) streams within simulations does allow us to ensure independent responses at all design points, the use of *common random numbers* (CRNs) is increasingly popular, as it is helpful in the debugging phase of scenario development.

In Gill (2019) I used a numerical experiment with a stochastic simulation (JFORCE, see Au et al. (2018)) to illustrate how the precision (variance) associated with linear regression coefficients can differ if iid is assumed, and the subsequent possibility of making incorrect inferences such as false negatives (declaring a factor as unimportant when it isn't) as a result. However, that paper necessarily skipped over much of the underpinning statistical and mathematical derivations. It also alluded to, but did not expand upon, an alternative motive for employing CRNs, as a *variance reduction technique* (VRT) similar to the use of blocking in physical experiments, which practitioners may be unaware of.

The intent of this paper is to fill those gaps. Purely for ease of exposition purposes, it will use very simple linear regressions (one or two factors) to enable derivations of various regression coefficient confidence interval constructions, as well as to illustrate how the assignment of PRNs might be exploited to our (analytical) advantage. Doing so will clearly illustrate how heteroscedasticity and dependence can influence linear regression analysis. Seminal references, though perhaps less well-known nowadays, on how these simple illustrations generalise to more practical multiple linear regression problems will then be described.

Construction of *confidence intervals* (CIs) for linear regression modelling is a key element of simulation analytics, to test the statistical significance of the influential factors and to bound their magnitude. However, too often the simplifying assumptions of iid residuals are used in statistical texts and/or software. This paper hopes to persuade the reader that such assumptions need not be made, and by reintroducing the work of Scheffé (1959) provide the mathematical background to procedures in the general case. In particular, the notion that independence of the simulation response to a designed experiment is a virtue, is hopefully dispelled. Indeed, the unique ability to control simulation's randomness should instead be viewed as an opportunity, to increase the precision of the linear regression CIs. The seminal work of Schruben & Margolin (1978) provides the optimal strategy of assigning the PRN streams for this goal.

Future research will focus on two extensions. First, while factorial designs are known to be optimal for multivariate linear regression when iid assumptions are met, the presence of heteroscedasticity has been shown to require the search for an alternative optimal design (Atkinson & Cook 1995). Second, for simulation responses that are not continuous (e.g., binary or count), *generalised linear regression* is used, and it is not clear that the assignment strategy of Schruben & Margolin (1978) will automatically apply. I hope to share the outcomes from this research at MODSIM 2023.

Keywords: *Multivariate linear regression, Variance reduction, Heteroscedasticity, Confidence ellipsoid*

1 INTRODUCTION

Simple linear regression (SLR) involves one factor x sampled at two or more levels. Suppose, however, that we take only two observations. Without loss of generality let $x_1 = +1$ and $x_2 = -1$, and let y_1 and y_2 denote the respective response values. Using the most common fitting criteria of *Ordinary Least Squares* (OLS), then the estimators for the two regression coefficients of the line that passes through these points ($\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$) are $\hat{\beta}_0 = \frac{1}{2}(y_1 + y_2)$ and $\hat{\beta}_1 = \frac{1}{2}(y_1 - y_2)$. To obtain *confidence intervals* (CIs) for the true β_0 and β_1 we need to treat $\hat{\beta}_0$ and $\hat{\beta}_1$ as *random variables* (RVs) and calculate both their mean and variance. Now:

$$E[\hat{\beta}_0] = \frac{1}{2}(\beta_0 + \beta_1 + E[\epsilon_1] + \beta_0 - \beta_1 + E[\epsilon_2]) = \beta_0 \text{ as } E[\epsilon_i] = 0 \tag{1}$$

and where $\epsilon_i = y_i - (\beta_0 + \beta_1 x_i)$. Likewise for $\hat{\beta}_1$, which confirms that these estimators are *unbiased*. Concerning the variance, if y_1 and y_2 are *independent* then:

$$var(\hat{\beta}_0) = var(\frac{y_1}{2}) + var(\frac{y_2}{2}) = \frac{1}{4}(\sigma_1^2 + \sigma_2^2) \tag{2}$$

where $var(y_i) = var(\epsilon_i) = \sigma_i^2$. We can likewise trivially show that $var(\hat{\beta}_1) = var(\hat{\beta}_0)$. However, the two estimators' covariance:

$$cov(\hat{\beta}_0, \hat{\beta}_1) = cov(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}) = \frac{\sigma_1^2 - \sigma_2^2}{4} \neq 0 \text{ if } \sigma_1^2 \neq \sigma_2^2, \tag{3}$$

so *heteroscedasticity* (unequal variances) induces *correlated* OLS estimators, which falsifies the notion that independence and an orthogonal design (which $x_1 = +1$, $x_2 = -1$ is) are sufficient to analyse $\hat{\beta}_0$ and $\hat{\beta}_1$ independently. If y_1 and y_2 are also not independent then:

$$var(\hat{\beta}_0) = \frac{1}{4}var(y_1) + \frac{1}{4}var(y_2) + \frac{1}{2}cov(y_1, y_2) = \frac{\sigma_1^2 + \sigma_2^2 + 2\sigma_{12}}{4} \tag{4}$$

$$var(\hat{\beta}_1) = \frac{1}{4}var(y_1) + \frac{1}{4}var(y_2) - \frac{1}{2}cov(y_1, y_2) = \frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}{4} \tag{5}$$

where $cov(y_i, y_j) = cov(\epsilon_i, \epsilon_j) = \sigma_{ij} \neq 0$ ($cov(\hat{\beta}_0, \hat{\beta}_1)$ remains unchanged), so *dependency* (non-independent data) induces *unequal standard errors* (square root of the variance) in OLS estimators. This example therefore illustrates nicely an interesting duality, in that violation of the identically distributed data assumption causes dependent OLS estimators, and violation of the independent data assumption causes non-identically distributed OLS estimators.

The remainder of the paper is organised as follows. Section 2 derives mathematically various CIs for SLR, depending on the assumptions made. Motivated by observing unequal estimator variances, Section 3 then uses linear regression with two factors to illustrate how the user's control over the implementation of randomness in simulation might be exploited to improve the efficiency of statistical analyses. Section 4 briefly (re)introduces key elements of seminal works by Scheffé (1959) and Schruben & Margolin (1978) which generalise these illustrations, before Section 5 concludes with potential research avenues.

2 SIMPLE LINEAR REGRESSION AND CONFIDENCE INTERVAL CONSTRUCTION

Given possible violations of the iid assumptions, how might OLS CIs be constructed? For concreteness, and for ease of exposition, assume that we know that $\sigma_1^2 = 1, \sigma_2^2 = 2, \sigma_{12} = 0.5$ so $var(\hat{\beta}_0) = 1, var(\hat{\beta}_1) = 0.5$ and $cov(\hat{\beta}_0, \hat{\beta}_1) = -0.25$. If we simply ignore the fact the estimators are correlated (i.e., assumed $\hat{\beta}_0$ and $\hat{\beta}_1$ are now independent), the well-known *individual* CIs for β_i would be given by $\hat{\beta}_i \pm \sqrt{var(\hat{\beta}_i)} \times z_{\alpha/2}$ would apply, where $z_{\alpha/2}$ is the *critical value* from the *Normal distribution* and $1 - \alpha$ is the *confidence level*, so that:

$$\textit{CIs: Identically Distributed}: \hat{\beta}_0 \pm 1.000 \times z_{\alpha/2} \text{ and } \hat{\beta}_1 \pm 0.707 \times z_{\alpha/2}. \tag{6}$$

If we further assumed independence, then since $var(\hat{\beta}_i) = \frac{1}{4}(\sigma_1^2 + \sigma_2^2)$, the result become:

$$\textit{CIs: iid}: \hat{\beta}_0 \pm 0.866 \times z_{\alpha/2} \text{ and } \hat{\beta}_1 \pm 0.866 \times z_{\alpha/2}. \tag{7}$$

However, the following analysis illustrates that one need not make these assumptions. Consider new RVs given by $\hat{\mathbf{z}} = M(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ where $M = \left(\begin{smallmatrix} +0.608 & +1.467 \\ -0.879 & +0.364 \end{smallmatrix}\right)$ (this choice of M will be explained later) so that:

$$\hat{z_0} = +0.608(\hat{\beta_0} - \beta_0) + 1.467(\hat{\beta_1} - \beta_1) \quad \hat{z_1} = -0.879(\hat{\beta_0} - \beta_0) + 0.364(\hat{\beta_1} - \beta_1).$$

It is easy to show that $E[\hat{z_j}] = 0 \; j = 0, 1$ (unbiased), but now:

$$\begin{aligned} var(\hat{z_0}) &= 0.608^2 var(\hat{\beta_0} - \beta_0) + 1.467^2 var(\hat{\beta_1} - \beta_1) + 2 \times 0.608 \times 1.467 cov(\hat{\beta_0} - \beta_0, \hat{\beta_1} - \beta_1) \\ var(\hat{z_1}) &= (-0.879)^2 var(\hat{\beta_0} - \beta_0) + 0.364^2 var(\hat{\beta_1} - \beta_1) + 2 \times (-0.879) \times 0.364 cov(\hat{\beta_0} - \beta_0, \hat{\beta_1} - \beta_1) \end{aligned}$$

and since $var(X + a) = var(X)$ and $cov(X + a, Y + b) = cov(X, Y)$ we find:

$$\begin{aligned} var(\hat{z_0}) &= 0.608^2 \times 1 + 1.467^2 \times 0.5 + 2 \times 0.608 \times 1.467 \times (-0.25) = 1 \\ var(\hat{z_1}) &= (-0.879)^2 \times 1 + 0.364^2 \times 0.5 + 2 \times (-0.879) \times 0.364 \times (-0.25) = 1 \\ cov(\hat{z_0}, \hat{z_1}) &= \ldots = 0. \end{aligned}$$

So $\hat{z_0}$ and $\hat{z_1}$ are independent, have zero mean and unit variance, and therefore individual CIs are given in the usual way $P(-z_{\alpha/2} \leq \hat{z_j} \leq z_{\alpha/2}) = 1 - \alpha \quad j = 0, 1$. In order to obtain CIs for β_0 and β_1 transforming the limits $\hat{z_j} = \pm z_{\alpha/2} \; j = 0, 1$ to the lines:

$$\hat{\beta_1} - \beta_1 = \frac{-0.608}{+1.467}(\hat{\beta_0} - \beta_0) \pm \frac{z_{\alpha/2}}{+1.467} \quad \text{and} \quad \hat{\beta_1} - \beta_1 = \frac{+0.879}{+0.364}(\hat{\beta_0} - \beta_0) \pm \frac{z_{\alpha/2}}{+0.364}$$

indicates that the *joint confidence region* (JCR) for β_0 and β_1 comprises a *rotated rectangle*. For *individual* CIs, the simplest construction would be to take the extent of the axes that lie within this JCR:

$$\hat{\beta_0} - \beta_0 \pm \frac{z_{\alpha/2}}{max(|0.608|, |-0.879|)} \quad \text{and} \quad \hat{\beta_1} - \beta_1 \pm \frac{z_{\alpha/2}}{max(|1.467|, |0.364|)}$$

which results in:

CIs: Rotated Rectangle: $\hat{\beta_0} \pm 1.14 \times z_{\alpha/2}$ and $\hat{\beta_1} \pm 0.68 \times z_{\alpha/2}$. (8)

2.1 Joint Confidence Regions

While (8) is certainly one way of constructing CIs without having to make iid assumptions, an issue is that if $P(-z_{\alpha/2} \leq X_j \leq z_{\alpha/2}) = 1 - \alpha, \quad j = 0, 1$, it is only true that $P\left(\bigcap_{j=0}^{1} \{-z_{\alpha/2} \leq X_j \leq z_{\alpha/2}\}\right) < 1 - \alpha$, and can be as small as $1 - 2\alpha$ (or for p factors as little as $1 - (p + 1)\alpha$), a result known as the Bonferroni inequality (see Law (2007)). However, if we exploit the fact $\hat{z_0}^2 + \hat{z_1}^2 \sim \chi^2_2$, so that $P\left(\hat{z_0}^2 + \hat{z_1}^2 \leq \chi^2_{2,\alpha}\right) = 1 - \alpha$ (exact, not an upper bound, and where $\chi^2_{2,\alpha}$ is the critical value from the *Chi-Squared* distribution with 2 degrees of freedom), since $\hat{\mathbf{z}} = M(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $\hat{z_0}^2 + \hat{z_1}^2 = \hat{\mathbf{z}}'\hat{\mathbf{z}}$, we find that $P\left((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \left((M'M)^{-1}\right)^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq \chi^2_{2,\alpha}\right) = 1 - \alpha$ (where M was defined after (7)). Finally, with the *eigendecompostion*:

$$\Sigma_{\hat{\boldsymbol{\beta}}} = VUV' = \begin{pmatrix} +0.383 & -0.924 \\ +0.924 & +0.383 \end{pmatrix} \begin{pmatrix} 0.396 & 0 \\ 0 & 1.104 \end{pmatrix} \begin{pmatrix} +0.383 & +0.924 \\ -0.924 & +0.383 \end{pmatrix} \tag{9}$$

we see that choosing:

$$M = U^{-1/2}V' = \begin{pmatrix} +0.608 & +1.467 \\ -0.879 & +0.364 \end{pmatrix} \rightarrow (M'M)^{-1} = \begin{pmatrix} 1 & -0.25 \\ -0.25 & 0.5 \end{pmatrix} = \Sigma_{\hat{\boldsymbol{\beta}}} \tag{10}$$

which results in:

$$P\left((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \Sigma^{-1}_{\hat{\boldsymbol{\beta}}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq \chi^2_{2,\alpha}\right) = 1 - \alpha. \tag{11}$$

This JCR represents a *rotated ellipse* (the eigenvectors define the rotation and the eigenvalues define the semi-major and minor length). In fact, for SLR we can perform the eigendecomposition analytically:

$$\lambda_{1,2} = \frac{\sigma_1^2 + \sigma_2^2 \pm \sqrt{\left(\sigma_1^2 - \sigma_2^2\right)^2 + 4\sigma_{12}^2}}{4} \quad \vec{v}_{1,2} = \begin{pmatrix} 2\sigma_{12} \pm \sqrt{\left(\sigma_1^2 - \sigma_2^2\right)^2 + 4\sigma_{12}^2} \\ \sigma_1^2 - \sigma_2^2 \end{pmatrix} \tag{12}$$

so that the angle of rotation associated with the new coordinate axes is given by:

$$\theta = \arctan\left(\frac{\sigma_1^2 - \sigma_2^2}{2\sigma_{12} \pm \sqrt{\left(\sigma_1^2 - \sigma_2^2\right)^2 + 4\sigma_{12}^2}}\right). \tag{13}$$

We can see clearly how the three elements of variability: variance ($\sigma_1^2 + \sigma_2^2$); heteroscedasticity ($\sigma_1^2 - \sigma_2^2$); and correlation (σ_{12}), affect this ellipse. First, we note that only the latter two affect the angle of rotation: and if there is no heteroscedasticity, then $\theta = \arctan(0) = 0$ (no rotation), while if there is no correlation, then $\theta = \arctan(1) = \pi/4$ (45° rotation). For our particular example, $\sigma_1^2 = 1, \sigma_2^2 = 2, \sigma_{12} = 0.5$, we find $\lambda_{1,2} = (3 \pm \sqrt{2})/4$, $\vec{v}_{1,2} = \left(1 \pm \sqrt{2}, -1\right)'$ and $\theta = -22.5°$. For the individual CIs, ideally we want the *projections* of this ellipse onto the $(\hat{\beta}_0 - \beta_0, \hat{\beta}_1 - \beta_1)$ axes. Since this occurs when $d\beta_1/d\beta_0 = 0$ and ∞ and recalling that:

$$\hat{z_0} = +0.608(\hat{\beta}_0 - \beta_0) + 1.467(\hat{\beta}_1 - \beta_1) \text{ and } \hat{z_1} = -0.879(\hat{\beta}_0 - \beta_0) + 0.364(\hat{\beta}_1 - \beta_1) \tag{14}$$

we therefore have:

$$\hat{z_0}^2 + \hat{z_1}^2 = 1.143(\hat{\beta}_0 - \beta_0)^2 + 1.143(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1) + 2.286(\hat{\beta}_1 - \beta_1)^2. \tag{15}$$

Implicitly differentiating $\hat{z_0}^2 + \hat{z_1}^2 = \chi^2_{2,\alpha}$ yields:

$$\frac{d\beta_1}{d\beta_0} = [-2.286(\hat{\beta}_0 - \beta_0) + 1.143(\hat{\beta}_1 - \beta_1)][4.572(\hat{\beta}_1 - \beta_1) + 1.143(\hat{\beta}_0 - \beta_0)]^{-1} \tag{16}$$

so $d\beta_1/d\beta_0 = 0$ when $\hat{\beta}_0 - \beta_0 = -0.5(\hat{\beta}_1 - \beta_1)$ and the intercepts are $\hat{\beta}_1 - \beta_1 = \pm\sqrt{0.5\chi^2_{2,\alpha}}$ (note $0.5 = var(\hat{\beta}_1)$). Likewise, $d\beta_1/d\beta_0 = \infty$ when $\hat{\beta}_1 - \beta_1 = -0.25(\hat{\beta}_0 - \beta_0)$ and the intercepts become $\hat{\beta}_0 - \beta_0 = \pm\sqrt{\chi^2_{2,\alpha}}$ (note $1 = var(\hat{\beta}_0)$), so the individual CIs from this approach become:

$$\textit{CIs with no iid Assumptions}\text{: } \hat{\beta}_0 \pm 1.000 \times \sqrt{\chi^2_{2,\alpha}} \text{ and } \hat{\beta}_1 \pm 0.707 \times \sqrt{\chi^2_{2,\alpha}}. \tag{17}$$

In summary, we see there are several possible CIs, given by (6-8, 17), depending on the assumptions made and the construction method. The most commonly made assumes iid residuals (7), but it is the latter (17) that is actually the preferred procedure, as it does not impose additional assumptions. It should be noted that these examples were for ease of exposition, and in practice σ_1^2, σ_2^2 and σ_{12} would be estimated from sample data, and that the Z and χ^2 distributions would be replaced by the t and F distributions. Fortunately, and as indicated by the matrix formulations of (9-11), this also generalises for *multiple linear regression* with p factors, as will be outlined in Section 4.

3 MULTIPLE LINEAR REGRESSION AND VARIANCE REDUCTION

3.1 Simulation Randomness Control

For SLR the regression coefficient estimates variances ($var(\hat{\beta}_0) = 1, var(\hat{\beta}_1) = 0.5$) were unequal, and examining (4) this was due to the dependency between responses at the (two) design points. There, a positive dependency (σ_{12}) increased the variance of the estimate of the mean ($var(\hat{\beta}_0) = 1$) while simultaneously decreasing the variance of the estimate of the (single) factor effect ($var(\hat{\beta}_1) = 0.5$). Usually, the factor effect(s) are of primary importance, so this result would seem fortuitous, as we would have a smaller CI than if there were no dependency between the responses (i.e., independence). Conversely, if there was a negative dependency ($\sigma_{12} < 0$), the variance of the estimate of the mean would decrease while that of the estimate of the (single) factor effect would go up. In simulations, the dependency (or independence) and direction (positive or negative) of the responses at the design points are determined by how randomness is implemented. At each design point, a stream of so-called pseudo-random numbers (PRNs) must be made available so that the stochastic elements within the simulation can be executed. To ensure (the assumption of) independence, non-overlapping (random) PRN streams must be assigned to each of the design points.

But as we have just seen from the SLR example, independence is not necessarily a virtue, and we can do better (in the sense of a tighter CI for the factor effect) if we could induce a positive (but not negative) correlation

at our two design points. This can be affected by using the same PRN streams at these design points, and this assignment is called common random numbers (CRNs). While not the focus of this paper, it should be noted that the degree of correlation achieved depends on how synchronised the simulation runs are, where ideally each source of random variation is given its own separately-seeded stream.

3.2 Common Random Numbers

The benefits of CRN can extend beyond SLR. For ease of exposition, let's consider the simplest multiple linear regression, that is with two factors x_1 and x_2 and the following linear regression model:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i} + \epsilon_i(x_{1i}, x_{2i}) \quad i = 1, \ldots 4 \quad \boldsymbol{\epsilon} \sim \mathbf{RV}(\mathbf{0}, \Sigma_{\boldsymbol{\epsilon}}). \tag{18}$$

where $\Sigma_{\boldsymbol{\epsilon}}$ is the covariance matrix of the regression residuals. Fitting general linear regression with OLS, it is well known (see Kleijnen (2015) for example) that $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$ and the covariance matrix for $\hat{\boldsymbol{\beta}}$ is given by:

$$\Sigma_{\hat{\boldsymbol{\beta}}} = (X'X)^{-1}X'\Sigma_{\boldsymbol{\epsilon}}X(X'X)^{-1}, \tag{19}$$

where X is the matrix of n design points. If X is orthogonal, then $X'X = nI$ and $\hat{\boldsymbol{\beta}} = X'\mathbf{y}/n$ with $\Sigma_{\hat{\boldsymbol{\beta}}} = X'\Sigma_{\boldsymbol{\epsilon}}X/(n^2)$. For concreteness, suppose we use the following (orthogonal) design matrix, and for ease of exposition assume homoscedasticity (equal variances at the design points), so that:

$$X = \begin{pmatrix} +1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & -1 & -1 & +1 \end{pmatrix} \quad \Sigma_{\boldsymbol{\epsilon}} = \sigma^2 \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{12} & 1 & \rho_{23} & \rho_{24} \\ \rho_{13} & \rho_{23} & 1 & \rho_{34} \\ \rho_{14} & \rho_{24} & \rho_{34} & 1 \end{pmatrix} \tag{20}$$

where σ^2 is the presumed constant variance and ρ_{ij} is the correlation between $y_i = y(\mathbf{x}_i)$ and $y_j = y(\mathbf{x}_j)$. For this two factor general linear regression it is not too difficult to algebraically express the diagonal entries of $\Sigma_{\hat{\boldsymbol{\beta}}}$ which are the variances of the regression coefficient estimates:

$$var(\hat{\beta}_0) = \frac{\sigma^2}{4}\left(1 + \frac{\rho_{12} + \rho_{13} + \rho_{14} + \rho_{23} + \rho_{24} + \rho_{34}}{2}\right) \tag{21}$$

$$var(\hat{\beta}_1) = \frac{\sigma^2}{4}\left(1 + \frac{\rho_{12} - \rho_{13} - \rho_{14} - \rho_{23} - \rho_{24} + \rho_{34}}{2}\right) \tag{22}$$

$$var(\hat{\beta}_2) = \frac{\sigma^2}{4}\left(1 + \frac{-\rho_{12} + \rho_{13} - \rho_{14} - \rho_{23} + \rho_{24} - \rho_{34}}{2}\right) \tag{23}$$

$$var(\hat{\beta}_{12}) = \frac{\sigma^2}{4}\left(1 + \frac{-\rho_{12} - \rho_{13} + \rho_{14} + \rho_{23} - \rho_{24} - \rho_{34}}{2}\right). \tag{24}$$

Here, the situation is not as straight-forward as in SLR. There are six correlations (ρ_{ij}) and they occur in the four variances with differing signs. However, if we used CRNs at the four design points, and if we assume for ease of illustration that the induced positive correlations each had the same magnitude (so $\rho_{ij} = \rho_+ > 0 \, \forall i, j$) then (21-24) indicate that the variance of the estimators of the mean and factor effects are:

$$var(\hat{\beta}_0) = \frac{\sigma^2}{4}(1 + 3\rho_+), \quad var(\hat{\beta}_k) = \frac{\sigma^2}{4}(1 - \rho_+) \; k = 1, 2, \quad var(\hat{\beta}_{12}) = \frac{\sigma^2}{4}(1 - \rho_+) \tag{25}$$

which would, again, be fortuitous if we were interested in the two main effects and their two-way interaction.

3.3 Antithetic Random Numbers

But suppose we were not interested in the two-way interaction, or assumed (or somehow knew) it wasn't a significant effect in the simulation. This is sometimes done in the early stages of simulation analyses where there are many, potentially significant, factors and the goal is to identify, or screen, these efficiently. How might we control the simulation randomness in our favour for this situation?

A critical observation that will assist us is that if we add the variances of the four estimators in (21-24), all of the correlations/covariances cancel out and we are left with simply the average variance σ^2 (note that this

was also true for our SLR example). This reflects a general principle of *variance invariance* or an example of the *no free lunch theorem*. That is, the inherent variability of the simulation is reflected (in totality) in the variability of the estimators in linear regression models.

However, rather than seeing this as a constraint, it actually provides the motivation for the screening strategy. We note that for our simple, general linear regression model, the use of CRNs apportioned more of the variance σ^2 onto the estimator of the mean $\hat{\beta}_0$, due to all correlations in (21) appearing with positive coefficients (and thus less variance on the other three estimators). If, however, we are happy with sacrificing precision on the two-way interaction (and consequently gain precision on the other three estimators-in essence, achieving the so-called *variance swindle* or variance reallocation), then the strategy would require generating positive correlations ρ_{14} and ρ_{23} and negative correlations elsewhere. To induce a negative correlation in the simulation response between two design points, for each uniform random number in the PRN stream assigned to one design point, we use the complement (i.e., the value 1 minus that random number) for the other design point, thus forming a stream of *antithetic* random numbers (ARNs).

Now, for the simple general linear model, $\rho_{14} > 0$ if $\mathbf{x}_1$ and $\mathbf{x}_4$ share the same PRN stream (i.e., CRNs) and $\rho_{23} > 0$ if $\mathbf{x}_2$ and $\mathbf{x}_3$ share the same PRN stream (i.e., CRNs). However, if we made both pairs (i.e., all design points) share the same PRN stream, then we would not induce the required negative correlations. But, if we make $\mathbf{x}_1$ and $\mathbf{x}_4$ share the same PRN stream, and $\mathbf{x}_2$ and $\mathbf{x}_3$ share the same ARN stream (this ARN stream being the complement to the PRN stream), then we would induce both the positive and negative correlations required. Assuming, as before, $\rho_+ > 0$ and similarly $\rho_- > 0$ (i.e., equal magnitudes of induced correlations) then (21) indicate that the variance of the estimators of the mean and (main) factor effects are:

$$var(\hat{\beta}_0) = \frac{\sigma^2}{4}(1 + \rho_+ - 2\rho_-), \quad var(\hat{\beta}_k) = \frac{\sigma^2}{4}(1 - \rho_+) \; k = 1, 2 \tag{26}$$

where we now have greater precision in the estimator for the mean (along with the previous precision on the main effects estimators).

In summary, we see that contrary to conventional wisdom, independence of the response variable to a designed experiment is not necessarily a virtue, and in simulation experiments the unique ability to control randomness provides the basis for a technique to potentially reduce the variance associated with the linear regression estimators of factors of interest, thereby increasing their precision. The simplest general linear regression (i.e., with two factors) illustrated how the assignment of PRN streams (both common and antithetic) to the design points can be made to greatest (analytical) effect. As will be outlined in the next section, this result also generalises for multivariate linear regression with p factors.

4 MULTIPLE LINEAR REGRESSION GENERALISATIONS

The occasional use of matrix formulations in discussing SLR and the appropriate CI construction makes it relatively easy to convince the reader that the results illustrated in Section 2 naturally carry over in the general case of linear regression with $p > 1$ factors. Here, $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and the OLS estimator is $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$.

If $\hat{\mathbf{z}} = M(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ then $\Sigma_{\hat{\mathbf{z}}} = \Sigma_{M\hat{\boldsymbol{\beta}} - M\boldsymbol{\beta}} = \Sigma_{M\hat{\boldsymbol{\beta}}} = M\Sigma_{\hat{\boldsymbol{\beta}}}M'$ since $M\boldsymbol{\beta}$ is constant. With the eigendecomposition $\Sigma_{\hat{\boldsymbol{\beta}}} = VUV'$ and choosing $M = U^{-1/2}V'$ this becomes:

$$\Sigma_{\hat{\mathbf{z}}} = U^{-1/2}V'VUV'VU^{-1/2} = U^{-1/2}IUIU^{-1/2} = I \tag{27}$$

since V is an orthogonal matrix (hence $V' = V^{-1}$). Thus, the elements of $\hat{\mathbf{z}}$ are independent, unit-variance distributed RVs for which it is well-know that $\hat{\mathbf{z}}'\hat{\mathbf{z}} \sim \chi^2_{p+1}$, so that $P\left(\hat{\mathbf{z}}'\hat{\mathbf{z}} \leq \chi^2_{p+1,\alpha}\right) = 1 - \alpha$. Finally, $\hat{\mathbf{z}}'\hat{\mathbf{z}} = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'M'M(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and with $M = U^{-1/2}V'$ we have:

$$M'M = VU^{-1/2}U^{-1/2}V' = VU^{-1}V' = ((V')^{-1}UV^{-1})^{-1} = (VUV')^{-1} = \Sigma^{-1}_{\hat{\boldsymbol{\beta}}} \tag{28}$$

which results in the JCR (11) for the multivariate case.

The aspect that is not so easily seen as generalising is the projection of the ellipse onto the regression coefficient coordinate axes. For SLR, the derivative $d\beta_1/d\beta_0$ was used to project the ellipse tangents perpendicular to the axes, and it was noted for the specific numerical example that these projected intercepts only depended on $var(\hat{\beta}_i,\, i = 0, 1)$ and not their covariance $cov(\hat{\beta}_0, \hat{\beta}_1)$. But these results do indeed hold in the multivariate case, thanks largely to the method of Scheffé's F-projections as proved (either algebraically or geometrically)

in the classic, but perhaps largely forgotten text, *The Analysis of Variance* by (Scheffé 1959). Thus the resultant individual CIs are, in general, given by $\hat{\beta}_i = \pm\sqrt{var(\hat{\beta}_i)\chi^2_{p+1,\alpha}}, \quad i = 0, \ldots, p$, where $var(\hat{\beta}_i)$ are the diagonal entries of $\Sigma_{\hat{\boldsymbol{\beta}}} = (X'X)^{-1}X'\Sigma_{\boldsymbol{\epsilon}}X(X'X)^{-1}$.

For variance reduction (or perhaps more correctly *variance reallocation* due to the variance invariance principle), algebraically expanding the expressions for $var(\hat{\beta}_i)$ for a given n-point design matrix (X) but general residual covariance matrix ($\Sigma_{\boldsymbol{\epsilon}}$) is particularly cumbersome. Schruben & Margolin (1978) showed that exhaustive search for the simple, multiple linear regression (4 point design matrix) required examining 49 different $\Sigma_{\boldsymbol{\epsilon}}$ to find the optimal assignment of CRN and ARN of Section 3. However, from that analysis they observed that the critical consequence of that assignment was that it partitioned the design into two blocks that confounded the interaction term of the regression model with the PRN effects. An n-point design matrix $X = (1, X_*)$ admits *orthogonal blocking* into 2 blocks if there exists an $n \times 2$ *block incidence matrix* W such that $1'W$ is a vector of positive integers and $X'_*W = 0$. Many experimental designs, including the 2^{n-k} fractional factorial designs, admit orthogonal blocking. The optimal assignment rule determined by Schruben & Margolin (1978) was to use the same stream of PRNs in one block and the corresponding ARN stream in the other block, and the greatest benefit is had if both blocks are of the same size. For a more recent publication, see Chih (2013).

5 DISCUSSION

Construction of CIs for linear regression modelling is a key element of simulation analytics, to test the statistical significance of the influential factors and to bound their magnitude. However, too often the simplifying assumptions of iid residuals are used in statistical texts and/or software. This paper hopes to persuade the reader that such assumptions need not be made, and by reintroducing the work of Scheffé (1959) provides the mathematical background to procedures in the general case. In particular, the notion that independence of the simulation response to a designed experiment is a virtue, is hopefully dispelled. Indeed, the unique ability to control simulation's randomness should instead be viewed as an opportunity, to increase the precision of the linear regression CIs. The seminal work of Schruben & Margolin (1978) provides the optimal strategy of assigning the PRN streams for this goal.

Future research will focus on two extensions. First, while factorial designs are known to be optimal for multivariate linear regression when iid assumptions are met, the presence of heteroscedasticity has been shown to require the search for an alternative optimal design (Atkinson & Cook 1995). Second, for simulation responses that are not continuous (e.g., binary or count), *generalised linear regression* is used, and it is not clear that the assignment strategy of Schruben & Margolin (1978) will automatically apply. I hope to share the outcomes from this research at MODSIM 2023.

ACKNOWLEDGEMENT

The author thanks Dr Averill Law and Professor Jack Kleijnen for fruitful discussions and the two anonymous referees for their critical reviews which improved the quality of this paper.

REFERENCES

Atkinson, A. C. & Cook, R. D. (1995), 'D-optimum designs for heteroscedastic linear models', *Journal of the American Statistical Association* **90**(429), 204–212.

Au, T. A., Hoek, P. J. & Lo, E. H. S. (2018), Combat analysis of joint force options using agent-based simulation, *in* '2018 Military Communications and Information Systems Conference (MilCIS)', pp. 1–7.

Chih, M. (2013), 'A more accurate second-order polynomial metamodel using a pseudo-random number assignment strategy', *Journal of the Operational Research Society* **64**(2), 198–207.

Gill, A. (2019), Two common pitfalls applying design of experiments (and hopefully how to avoid them!), *in* 'Proceedings of MODSIM2019, 23rd International Conference on Modelling and Simulation'.

Kleijnen, J. (2015), *Design and Analysis of Simulation Experiments*, 2nd edn, Springer, New York, USA.

Law, A. (2007), *Simulation Modeling and Analysis*, 4th edn, McGraw-Hill, Boston, USA.

Montgomery, D. (2012), *Design and Analysis of Experiments, 8th Edition*, John Wiley & Sons, Incorporated.

Scheffé, H. (1959), *The analysis of Variance*, Wiley.

Schruben, L. W. & Margolin, B. H. (1978), 'Pseudorandom number assignment in statistically designed simulation and distribution sampling experiments', *Journal of the American Statistical Association* **73**(363), 504–520.