Supplementary materials for this article are available online. Please click the Technometrics link at http://pubs.amstat.org.

Leveraged Gauge R&R Studies

Ryan BROWNE, Jock MACKAY, and Stefan STEINER

Business and Industrial Statistics Research Group
Department of Statistics and Actuarial Sciences
University of Waterloo
Waterloo, N2L 3G1, Canada
(*rpbrowne@gmail.com*)

To assess measurement system variation, we propose an alternative to the standard gauge reproducibility and repeatability (GR&R) study. The new plan, called a leveraged GR&R Study, is conducted in two stages. In the baseline stage, we select a sample of parts that are measured once only each using a fixed number of operators. Then we deliberately select extreme parts for the second stage where each operator measures each selected part a number of times. We demonstrate the advantages of the leveraged over the standard plan by comparing the standard deviations of the estimators of the parameters of interest. For a fixed number of operators and total number of measurements, we recommend leveraged plans with a baseline size that is roughly half the total number of measurements. We also recommend that the number of parts selected for the second stage be set to an integer multiple of the number of operators and that each of these parts be measured two or three times by each operator. This article has supplementary material online.

KEY WORDS: Leveraging; Measurement system assessments; Repeatability; Reproducibility.

1. INTRODUCTION

Operators are often thought to be a substantial source of variability in a measurement system. Each operator is assumed to have a different mean when repeatedly measuring the same part so that there are relative biases among the operators. In this article, we consider extending previous work on leveraging by (Browne, MacKay, and Steiner 2009a, 2009b) to the assessment of a measurement system with multiple operators. We use the multiple operator terminology, but the extension is also applicable when there are multiple parallel gauges in the system with no operator effects.

The most common measurement assessment plan for a system which includes operators is called a gauge reproducibility and repeatability (GR&R) study as described in Automotive Industry Action Group (2003). With a GR&R, we can partition measurement variation into two sources: repeatability and reproducibility. Repeatability is the variation associated with repeatedly measuring the same part with the same operator. Reproducibility is the variation attributable to differences among operators.

To assess the repeatability and reproducibility of a measurement system, we repeatedly measure a number of parts with m different operators. For a measurement Y_j from the jth operator, $j = 1, \ldots, m$, on a randomly selected part, we adopt the model

$$Y_j = \mu_j + P + E, \tag{1}$$

where μ_j is the mean for operator j, P is a random variable whose distribution describes the possible true values of the part characteristic, and E represents the random measurement error. The random variables P and E are assumed to be independent and normal with means μ_p and 0 and standard deviations σ_p and σ_g, respectively. We can only estimate $\mu_j + \mu_p$ so we set $\mu_p = 0$.

The operator effects μ_j can be modeled as random or fixed. A random effects model describes the situation where the measurement system is used by a large number of operators and only a sample of these operators are used in the assessment study. Burdick, Borror, and Montgomery (2005), Montgomery and Runger (1993a), and Wheeler and Lyday (1984) use a random effects model in the analysis of a GR&R study. We recommend assuming fixed effects for operators when only a few operators use the measurement system in production. One argument in favor of fixed effects comes from Dolezal, Burdick, and Birch (1998). They noted that a GR&R study is required every time the production process changes and occasionally for routine measurement assessment. The time between assessments can be short, making it realistic that only a few operators use the measurement system during the time between studies. They also note that if a mixed model (operators with fixed effects) better represents the measurement system, using a random effects model (operators with random effects) will result in unnecessarily accounting for extra variability.

With operators having fixed effects, we need to define carefully the parameters of interest. Suppose we have m operators each with a fixed effect quantified by the mean μ_j in Equation (1). To define reproducibility, let

$$\sigma_o^2 = \frac{1}{m}\sum_{j=1}^{m}(\mu_j - \overline{\mu})^2, \tag{2}$$

where $\overline{\mu} = \sum(\mu_j/m)$. The parameter σ_o captures the variation due to differences among the m operator means (i.e., the effects

DOI 10.1198/TECH.2010.09037

of relative bias), but is not a variance in the usual sense. Next, we define

$$\sigma_{pg}^2 = \sigma_p^2 + \sigma_g^2, \tag{3}$$

$$\sigma_t^2 = \sigma_{pg}^2 + \sigma_o^2, \tag{4}$$

$$\sigma_{go}^2 = \sigma_o^2 + \sigma_g^2. \tag{5}$$

The parameter σ_{pg} is the variation seen in measurements made by any single operator on a sample of parts from the process. The parameter σ_t represents the total variation seen in the process if each operator measured the same proportion of parts in regular production. Finally, the parameter σ_{go} represents the total variation seen in the measurement of any particular part if each operator is used with the same intensity. If each operator has the same mean, then $\sigma_o = 0$ and we can interpret σ_{go} and σ_t defined in Equation (1) as standard deviations. Using manufacturing jargon, σ_{go} represents the overall measurement variability, σ_g the repeatability, and σ_o the reproducibility. We assume σ_g is the same for each operator and part. By adopting Equation (1), we also assume overall process stability, i.e., that all of the parameters are constant over the time needed to conduct the assessment.

To quantify the contribution of the measurement system to the total variation, we use the ratio $\gamma = \sigma_{go}/\sigma_t$ that corresponds to the usual performance measure in the random effects model. Note that $0 \le \gamma \le 1$ and the smaller the value of γ, the smaller is the contribution of the measurement system to the overall variation. The commonly adopted cut-off values (see Automotive Industry Action Group 2003) used to determine if a measurement system is acceptable are $\gamma = 0.10$ and $\gamma = 0.30$. Any reasonable system has $\gamma < 0.5$.

When trying to improve a measurement system, a metric that compares the reproducibility and repeatability is of interest. We use $\lambda = \sigma_o^2/\sigma_{go}^2$, which is the ratio of variation due to the relative bias among the operators and overall measurement variation as defined in Equation (5). Note that λ uses variances whereas γ is based on standard deviations. We choose this scale so that $0 \le \lambda \le 1$ and $\lambda = 0.5$ corresponds to equal contributions from the repeatability and reproducibility. We treat γ as the primary parameter while λ is of secondary interest since the main purpose of the assessment is to validate the measurement system as a whole.

The standard measurement system assessment plan (see Dolezal, Burdick, and Birch 1998; Automotive Industry Action Group 2003; Burdick, Borror, and Montgomery 2003) is to sample k parts at random from the process and then have each of the m operators measure each part n times for a total of $N = kmn$ measurements. We denote the standard plan by SP. In practice, it is common to use use two or three operators ($m = 2, 3$), each of which measure the same $k = 10$ parts two or three ($n = 2, 3$) times for a total of 40 to 90 measurements. The analysis of the data and the corresponding estimates and confidence intervals for γ and λ are based on analysis of variance (ANOVA) using a model with fixed operator effects (Burdick, Borror, and Montgomery 2005, section 6.9). Here we use maximum likelihood estimation and look at asymptotic standard errors of the estimators. These standard errors are directly related to the length of confidence intervals, assuming the asymptotic approximations apply.

In this article, we introduce a two-stage leveraged plan (LP) where we first select a baseline sample of parts that are measured only once. Then we deliberately select extreme parts from the baseline to re-measure in the second stage. We demonstrate the advantages of the LP over the SP by comparing the asymptotic standard deviations of the estimators for γ and λ. The article has the following format. In Section 2, we explain the leveraged plan in detail and present the likelihood and Fisher information for the parameters of interest. We also give an example to illustrate how to estimate these parameters and obtain approximate standard errors. In Section 3, we compare the efficiency of the LP and SP. In Section 4, we propose good designs for leveraged plans. Finally, in Section 5, we draw conclusions and discuss possible extensions of leveraging.

2. LEVERAGED PLAN

A leveraged measurement system assessment is conducted in two stages:

Stage 1. Sample $(b \times m)$ parts at random from the process to obtain a baseline. Then allocate b different parts to each of the m operators which measure the allocated parts once. In this stage, no part is measured twice. We denote the initial measured value from part i and operator j as y_{ij0} for $i = 1, \ldots, b$ and $j = 1, \ldots, m$.

Stage 2. From the baseline sample, select k parts using the observed measured values. We denote the k selected parts by R which is a subset of $\{(1, 1), (1, 2), \ldots, (b, m)\}$ with k elements. To improve the estimation of γ and λ, sample these parts so that the initial measurements are extreme relative to their operator average, i.e., choose parts with large values of $|y_{ij0} - \overline{y}_{\cdot j0}|$ where $\overline{y}_{\cdot j0} = \frac{1}{b}\sum_{i=1}^{b} y_{ij0}$. Then use the k selected parts in a standard plan. That is, each of the m operators measures each of the k parts n times to give the additional data $\{y_{ijlh},\ (i, j) \in R,\ l = 1, \ldots, m$, and $h = 1, \ldots, n\}$. Note that y_{ijlh} is the hth measured value by operator l on the ith part measured by operator j in the baseline.

For example, in a leveraged plan with $b = 10, m = 3, k = 6$, and $n = 2$, we first sample 30 parts at random from the process and allocate 10 parts to each of the three operators. The operators measure each of their 10 assigned parts once. Then, one possibility is to pick the parts with the minimum and maximum initial measurement from each operator. In Stage 2, these $k = 6$ parts are used in a standard plan where each operator measures each part twice ($n = 2$). This plan has a total of $10 \times 3 + 6 \times 3 \times 2 = 66$ measurements. In general, the total number of measurements in the leveraged plan is $N = m(b + nk)$.

In Stage 1 of the LP, we must select and measure the parts over a sufficiently long time that we get a meaningful estimate of σ_t from the baseline sample. In Stage 2, we recommend repeatedly measuring the parts over the range of conditions (time, environment, etc.) expected to capture the major sources of measurement variation. We can alter the plan in this stage to quantify these sources of variation individually if that is deemed advisable. In this article, we do not consider the effect of leveraging on these altered plans. It is also helpful to select parts with large and small initial measurements to check the linearity of the measurement system.

2.1 Likelihood

To obtain the likelihood for the leveraged plan, we start by finding the likelihood for the repeated measurements on a single part. For any part (i, j) measured by operator j in the baseline, the joint distribution of the initial measurement Y_{ij0} and the n repeated measurements from each of the m operators is

$$\begin{pmatrix} Y_{ij0} \\ Y_{ij11} \\ \vdots \\ Y_{ijmn} \end{pmatrix} \backsim \mathrm{N}\left(\begin{bmatrix} \mu_j \\ \mu_1 \\ \vdots \\ \mu_m \end{bmatrix}, \sigma_{pg}^2[(1-\rho)I_{mn+1} + \rho J_{mn+1}] \right), \quad (6)$$

where $\rho = \sigma_p^2/\sigma_{pg}^2$, I_q is an identity matrix with dimension q, and J_q is a square matrix of ones with dimension q.

The distribution of the repeated measurements $\{Y_{ij11}, Y_{ij12}, \ldots, Y_{ijmn}\}$ on a single part, conditional on the initial measurement $Y_{ij0} = y_{ij0}$, is

$$\begin{pmatrix} Y_{ij11} \\ \vdots \\ Y_{ijmn} \end{pmatrix} \Bigg| Y_{ij0} = y_{ij0} \backsim \mathrm{N}\left(\begin{bmatrix} \mu_1 + \rho(y_{ij0} - \mu_j) \\ \vdots \\ \mu_m + \rho(y_{ij0} - \mu_j) \end{bmatrix} \otimes \underline{1}_n, \Sigma = \sigma_{pg}^2(1-\rho)(I_{mn} + \rho J_{mn}) \right), \quad (7)$$

where the Kronecker product $\otimes$ stacks n copies of the vector μ in a single vector and $\underline{1}_q$ is a column vector of ones with q rows. It is important to note that this conditional distribution depends only on the value of the initial measurement and not on how the part was selected. To understand this result, imagine after the initial measurements are made, all bm parts in the baseline are re-measured n times by each operator. You are given the initial, but not the additional measurements. Then, for each part, the previous conditional distribution in Equation (7) will apply. Selecting any subset of parts does not change the conditional distribution as long as the selection is based on the given initial values. The order of making the additional measurements and selecting the parts to be re-measured does not matter.

The covariance matrix Σ in Equation (7) has a special form that allows us to obtain the following, using well known properties (Dillon and Goldstein 1984):

$$\Sigma^{-1} = \frac{1}{\sigma_{pg}^2(1-\rho)(1+mn\rho)}\{[1+(mn-1)\rho]I_{mn} - \rho J_{mn}\},$$
$$|\Sigma| = \sigma_{pg}^{2mn}(1-\rho)^{mn}(1+mn\rho).$$

Using these properties of Σ, we can write the conditional likelihood (conditional on y_{ij0}) for the repeated measurements on each selected part. It is convenient to reparametrize the likelihood in terms of $\underline{\mu} = (\mu_1, \mu_2, \ldots, \mu_m)^t$ and

$$\sigma_{pg}^2 = \sigma_p^2 + \sigma_g^2 \qquad \text{and} \qquad \rho = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_g^2}. \quad (8)$$

We assume in Equation (1) that the measurements for one part are independent of the measurements from another. The conditional likelihood for all k parts in Stage 2, each with nm measurements, is the product of the individual likelihoods. The conditional log-likelihood is

$$l_2(\underline{\mu}, \sigma_{pg}^2, \rho | y_{ij0}, (i,j) \in R) = \left[-\frac{mnk}{2}\log\sigma_{pg}^2 - \frac{mnk}{2}\log(1-\rho) - \frac{k}{2}\log(1+mn\rho) \right] - \frac{1}{2}\frac{1}{\sigma_{pg}^2(1-\rho)(1+mn\rho)} \times \{(1+mn\rho)G(\underline{\mu}) + mnH(\underline{\mu}, \rho)\}, \quad (9)$$

where

$$G(\underline{\mu}) = \sum_{(i,j)\in R}\sum_{l=1}^{m}\sum_{h=1}^{n}(z_{ijlh} - \overline{z}_{ij\cdot\cdot})^2,$$
$$H(\underline{\mu}, \rho) = \sum_{(i,j)\in R}[\overline{z}_{ij\cdot\cdot} - \rho z_{ij0}]^2,$$

with $z_{ij0} = y_{ij0} - \mu_j$, $z_{ijlh} = y_{ijlh} - \mu_l$, and $\overline{z}_{ij\cdot\cdot} = \frac{1}{mn} \times \sum_{l=1}^{m}\sum_{h=1}^{n} z_{ijlh}$.

We next find the likelihood of the baseline data. In the baseline, each of the m operators measures b different parts for total of $b \times m$ measurements. Assuming these parts are selected at random from the process, the marginal log-likelihood of the baseline data is

$$l_1(\underline{\mu}, \sigma_{pg}^2) = -\frac{bm}{2}\log\sigma_{pg}^2 - \frac{1}{2\sigma_{pg}^2}\left\{ \sum_{j=1}^{m}\sum_{i=1}^{b}(y_{ij0} - \overline{y}_{\cdot j0})^2 + b\sum_{j=1}^{m}(\overline{y}_{\cdot j0} - \mu_j)^2 \right\}, \quad (10)$$

where $\overline{y}_{\cdot j0} = \frac{1}{b}\sum_{i=1}^{b} y_{ij0}$ is the average of the baseline measurements for operator j. Finally, the (unconditional) log-likelihood for the LP is

$$l(\underline{\mu}, \sigma_{pg}^2, \rho) = l_1(\underline{\mu}, \sigma_{pg}^2) + l_2(\underline{\mu}, \sigma_{pg}^2, \rho | y_{ij0}, (i,j) \in R). \quad (11)$$

To get the maximum likelihood estimates (MLE's) of $\underline{\mu}, \sigma_{pg}^2$, and ρ, we maximize Equation (11). Solutions can be found numerically. Then to get the MLE's for γ and λ, we apply the appropriate transformations to give

$$\widehat{\gamma} = \sqrt{\frac{\widehat{\sigma}_o^2 + (1-\widehat{\rho})\widehat{\sigma}_{pg}^2}{\widehat{\sigma}_o^2 + \widehat{\sigma}_{pg}^2}} \qquad \text{and} \qquad \widehat{\lambda} = \frac{\widehat{\sigma}_o^2}{\widehat{\sigma}_o^2 + \widehat{\sigma}_g^2}, \quad (12)$$

where we substitute $\widehat{\underline{\mu}}$ into Equation (2) to estimate $\widehat{\sigma}_o^2$. R code to maximize the likelihood and generate approximate standard errors is available at *http://www.bisrg.uwaterloo.ca/*.

2.2 Fisher Information and Asymptotic Variances

The asymptotic variance–covariance matrix of the maximum likelihood estimators is the inverse of the Fisher information matrix. We obtain the Fisher information matrix by summing several pieces. For each piece, we calculate minus the second derivatives of the log-likelihood with respect to the parameters

and then take expectations. We omit the tedious calculations. The first piece, from the baseline likelihood, is

$$J_1(\underline{\mu}, \sigma_{pg}^2, \rho) = \begin{bmatrix} \frac{b}{\sigma_{pg}^2}\mathbf{I}_m & \underline{0}_m & \underline{0}_m \\ & \frac{1}{2}\frac{mb}{\sigma_{pg}^4} & 0 \\ & & 0 \end{bmatrix}, \tag{13}$$

where $\mathbf{I}_m$ is the identity matrix with dimension m and $\underline{0}_m$ is a column vector of m zeros. Since information matrices are symmetric, we do not show the values below the diagonal. In addition to the baseline information matrix, we have one matrix for every part selected to be repeatedly measured in Stage 2. Each matrix has the same form. Suppose we have part i with initial measurement y_{ij0} measured by operator j. Then the Fisher information is

$$J_{2ij}(\underline{\mu}, \sigma_{pg}^2, \rho) = \begin{bmatrix} \mathbf{M}_0 + \mathbf{M}_{(j)} & \underline{0}_m & \underline{v}_{(i,j)} \\ & \frac{1}{2}\frac{mn}{\sigma_{pg}^4} & -\frac{1}{2}\frac{mn\rho(mn+1)}{\sigma_{pg}^2(1-\rho)(1+mn\rho)} \\ & & \frac{1}{2}\frac{nm(nm+1)(nm\rho^2+1)}{(1+mn\rho)^2(1-\rho)^2} + \frac{mn(E[Z_{ij0}^2]-1)}{(1+mn\rho)(1-\rho)} \end{bmatrix}, \tag{14}$$

where

$$\begin{aligned} \mathbf{M}_0 &= n(1+mn\rho)\mathbf{I}_m - n^2\rho\mathbf{J}_m, \\ \mathbf{M}_{(j)} &= n\rho(m\rho-1)\mathbf{I}_{(j)} - n\rho\mathbf{J}_{(m)}, \\ \mathbf{I}_{(j)} &= \underline{e}_j\underline{e}_j^t, \qquad \mathbf{J}_{(j)} = \underline{1}_m\underline{e}_j^t + \underline{e}_j\underline{1}_m^t - \underline{e}_j\underline{e}_j^t, \\ \underline{v}_{ij} &= \frac{nE[Z_{ij0}]}{\sigma_{pg}(1-\rho)(1+mn\rho)}\underline{1}_m - \frac{nm\rho E[Z_{ij0}]}{\sigma_{pg}(1-\rho)(1+mn\rho)}\underline{e}_j, \\ Z_{ij0} &= (Y_{ij0}-\mu_j)/\sigma_{pg}. \end{aligned} \tag{15}$$

Note that $\underline{e}_j$ is a column vector of length m with all zeros except for a single one at the jth position and $\underline{1}_m$ is a column vector of m ones.

To get the complete Fisher information, we add the baseline Fisher information and the Fisher information matrices over all parts used in Stage 2. We get

$$J(\underline{\mu}, \sigma_{pg}^2, \rho) = J_1(\underline{\mu}, \sigma_{pg}^2, \rho) + \sum_{(i,j)\in R} J_{2ij}(\underline{\mu}, \sigma_{pg}^2, \rho). \tag{16}$$

The second term in Equation (16) is a function of S_j, the sum of $E[Z_{ij0}]$ over all parts selected for Stage 2 and measured by operator j in the baseline, $j = 1, \ldots, m$, and $SS = \sum_{(i,j)\in R} E[Z_{ij0}^2]$. At this point, it is important to note that these expected values depend on how we select the second stage parts. The likelihood in Equation (11) depends on the initial measured values of the parts selected for Stage 2; the information in Equation (16) depends on how those parts are selected. Note that as SS increases (i.e., we select parts that are more extreme for Stage 2) so does the lower-right diagonal entry of the information matrix.

The Fisher information for $(\underline{\mu}, \lambda, \gamma)$ can be written in terms of $J(\underline{\mu}, \sigma_{pg}^2, \rho)$ and the matrix $\mathbf{D}$, which is the Jacobian of the transformation from $(\underline{\mu}, \sigma_{pg}^2, \rho)$ to $(\underline{\mu}, \lambda, \gamma)$. We have

$$\mathbf{D} = \begin{bmatrix} I_m & -\frac{2}{m}\frac{\lambda\gamma^2-1}{\lambda\gamma^2}(\underline{\mu} - \overline{\mu}\underline{1}_m) & \underline{0}_m \\ \underline{0}_m^t & -\frac{\sigma_a^2}{\lambda^2\gamma^2} & -\frac{2\sigma_a^2}{\lambda\gamma^3} \\ \underline{0}_m^t & \frac{\gamma^2(1-\gamma^2)}{(\lambda\gamma^2-1)^2} & \frac{2\gamma(\lambda-1)}{(\lambda\gamma^2-1)^2} \end{bmatrix}. \tag{17}$$

Then, the Fisher information for $(\underline{\mu}, \lambda, \gamma)$ is

$$J(\underline{\mu}, \lambda, \gamma) = \mathbf{D}J(\underline{\mu}, \sigma_{pg}^2, \rho)\mathbf{D}^t. \tag{18}$$

In the Appendix, we derive an expression for the asymptotic variance covariance matrix of the MLE's, the inverse of $J(\underline{\mu}, \lambda, \gamma)$, and show algebraically that the asymptotic variances of $\widehat{\gamma}$ and $\widehat{\lambda}$ decrease as SS increases with the other design parameters $S_1, \ldots, S_m$, n, m, k, and b fixed. That is, we get improved precision by selecting more extreme parts. If we select the k parts for Stage 2 at random, then $SS = k$. We can guarantee that $SS > k$ by selecting parts with relatively extreme initial values. Generally, we cannot change SS without changing $S_1, \ldots, S_m$. One common exception is the balanced LP where we chose an equal number of parts with large and small initial values for each operator so that $S_j = 0$ for each $j = 1, \ldots, m$.

To obtain an estimate of the asymptotic variance–covariance matrix for $(\widehat{\underline{\mu}}, \widehat{\lambda}, \widehat{\gamma})$ and hence approximate standard errors, we substitute the MLE's for the parameters in $J(\underline{\mu}, \lambda, \gamma)$ and then find the inverse numerically. We estimate the quantities $S_1, \ldots, S_m$ and SS using the corresponding sample quantities $z_{ij0} = (y_{ij0} - \widehat{\mu}_j)/\widehat{\sigma}_{pg}$. For a balanced design, we substitute $S_j = 0$ for each $j = 1, \ldots, m$.

2.3 An Example

To illustrate how to get estimates, standard errors, and check model assumptions for a leveraged plan, we consider an artificial example with $m = 3$ operators, and $b = 11, k = 3$, and $n = 3$. This means in Stage 1, 33 parts were sampled at random and distributed equally to the three operators. The baseline data are presented in Table 1. We selected parts numbered 4, 5, 11 (shown in bold) from operators 1, 2, 3, respectively, because of their extreme initial measurements. These three parts were then measured by each operator $n = 3$ times to yield the additional data given in Table 2.

Table 1. Baseline (Stage 1) data

	Baseline observations y_{ij0}											
j	$i=1$	2	3	4	5	6	7	8	9	10	11	$\overline{y}_{\cdot j0}$
1	−1.53	−1.35	−2.05	**2.12**	0.62	−2.03	−0.24	−1.22	−0.53	1.71	−0.35	−0.44
2	−1.34	0.36	−0.81	0.62	**−1.78**	1.11	−1.12	1.63	0.67	1.57	−0.25	0.06
3	1.58	0.55	1.56	0.80	1.81	−0.22	0.16	0.27	−0.53	−0.37	**1.93**	0.69

Table 2. Stage 2 data

	Repeated measurements for part (i, j)								
	(4, 1)			(5, 2)			(11, 3)		
	Repeats			Repeats			Repeats		
Operator	1	2	3	1	2	3	1	2	3
1	2.08	2.08	2.10	−1.91	−1.89	−1.92	1.80	1.77	1.77
2	2.24	2.19	2.22	−1.71	−1.81	−1.68	1.87	1.91	1.89
3	2.33	2.34	2.44	−1.67	−1.66	−1.62	2.02	1.97	1.95

The maximum likelihood estimates from Equation (11) for $(\mu_1, \mu_2, \mu_3, \sigma^2_{pg}, \rho)$ are $(-0.021, 0.113, 0.218, 1.425, 0.999)$. Using the transformations in Equation (12), the maximum likelihood estimates for (γ, λ) are $(0.087, 0.876)$. We can obtain standard errors by applying Equation (18) and substituting the MLE's. Here, the standard errors for γ and λ are 0.0120 and 0.0331, respectively. So, in this example, the measurement system variation is small with respect to the total variation in the process. An approximate 95% confidence interval for γ $(= \sigma_{go}/\sigma_t)$ is 0.087 ± 0.024 and since $\widehat{\lambda} = 0.876$, a substantial proportion of the measurement system variation is due to relative biases among the operators.

To check the model fit, we created a quantile–quantile (QQ) plot of the 33 baseline residuals defined as $y_{ij0} - \overline{y}_{\cdot j0}$. This plot is shown in the left panel of Figure 1. The right panel shows the baseline residuals by operator, which can be used to check the assumed constant standard deviation σ_g among the three operators. These two plots show no evidence to suspect the model.

To further check the model fit, we created a normal QQ plot of the 27 Stage 2 residuals defined as $y_{ijlh} - \overline{y}_{ijl\cdot}$ where $\overline{y}_{ijl\cdot} = \frac{1}{n}\sum_{h=1}^{n} y_{ijlh}$. This is shown in the left panel of Figure 2. The right panel shows the Stage 2 residuals plotted by operator that provides a second check of constant variance across the operators. Figures 1 and 2 give conflicting information regarding the assumption that σ_{go} is constant across operators and this illustrates how hard it is to compare standard deviations with a few degrees of freedom. Also, the residuals in Figure 1 include the dominant part-to-part variation which may completely mask any differences in σ_g among operators as suggested in Figure 2. We can also look informally for evidence of part-by-operator interaction by plotting the Stage 2 residuals by part with a separate plotting symbol for each operator. We do not include this plot here.

3. COMPARISON OF LEVERAGED TO STANDARD PLAN

To demonstrate the value of the leveraged plan we resort to simulation. We compare the LP to the SP when there are $m = 3$ operators and the total number of measurements is $N = 60$ and 90.

1. For $N = 60$, with $m = 3$ operators, we compare the following two plans
   - SP with $k = 10$ and $n = 2$ (a commonly used plan in practice)
   - LP with $b = 11$, $k = 3$, and $n = 3$ (as recommended in Section 4)
2. $N = 90$, with $m = 3$ we compare
   - SP with $k = 10$ and $n = 3$ (a commonly used plan in practice)
   - LP with $b = 18$, $k = 6$, and $n = 2$ (as recommended in Section 4).

To select the parts for Stage 2 of the LP with $k = 3$, we take the largest part from operator 1, the smallest from operator 2, and the largest from operator 3. When $k = 6$, we use the largest and smallest part from each operator.

To compare the SP and LP, we vary λ and γ, but fix $\sigma_t^2 = 1$ and the composition of μ_1, μ_2, μ_3. That is, for each simulation run we set $\sigma_o^2 = \gamma^2\lambda$ and then set $\mu_2 = 0$, $\mu_1 = -\mu_3$ and $\mu_3 = \sqrt{\frac{3}{2}\sigma_o^2}$. We need to fix the composition because there are an infinite number of possibilities for a particular value of σ_o^2. For example, $\underline{\mu}$ equal to $(-1.225, 0, 1.225)$ or $(-1.000, -0.366, 1.366)$ will yield the same value of σ_o^2. If σ_o^2 and $\overline{\mu} = \frac{1}{m}\sum_{i=1}^{m}\mu_j$ remain constant and we select the same number of parts from each operator in the baseline study, the asymptotic standard deviations of γ and λ, found in the inverse Fisher information, are the same no matter the composition of $\underline{\mu}$.

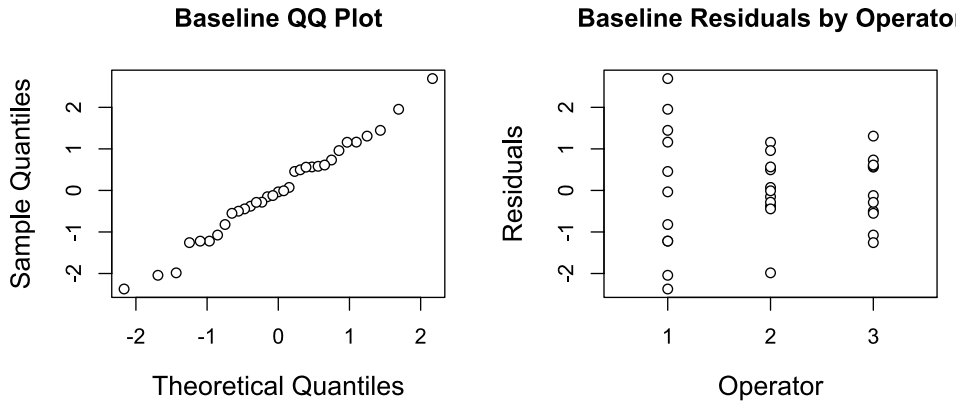

Figure 1. Model assessment using baseline data. The left panel shows a normal QQ plot of the baseline residuals $y_{ij0} - \overline{y}_{\cdot j0}$ and the right panel shows these residuals by operator.

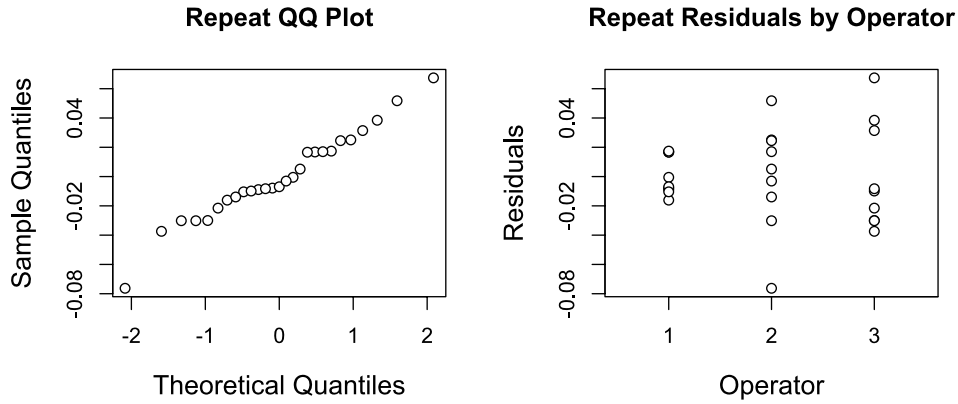

Figure 2. Model assessment using second stage data. The left panel shows a normal QQ plot of the second stage residuals $y_{ijlh} - \overline{y}_{ijl\cdot}$ and the right panel shows these residuals by operator.

In the comparison, we use MLE for the leveraged plan and ANOVA (Burdick, Borror, and Montgomery 2005) for the standard plan. We quantify the difference between the plans using the ratio of standard deviations of the estimators for γ and λ from the two plans. To calculate this ratio, we simulated 1000 repeats at a variety of γ and λ values. We restricted attention to $0 < \gamma < 0.5$ since it would be unusual to see a measurement system with $\gamma > 0.5$. Then, to create Figure 3, we smoothed the simulated values across the parameter space.

Figure 3 shows that with $N = 60$ the LP is substantially and uniformly more efficient than the SP for estimating the parameter of primary interest γ. If we have a good measurement system with $\gamma \leq 0.1$, the LP is 1.6 to 2 times more efficient in estimating γ. Conversely, the SP has a lower standard deviation for λ than the LP. The parameter λ is of secondary importance since if γ is small the measurement system is acceptable regardless of the λ value. Figure 4 compares the LP and SP when the total number of measurements is $N = 90$. Now the advantage of the LP over the SP for estimating γ is increased. The LP is two times more efficient than the SP almost everywhere. The advantage of the SP in estimating λ is reduced as well. We see similar results for other values of N and m. See the online supplemental material for more examples.

We show in the Appendix that the asymptotic variance of $\widehat{\lambda}$ decreases as we select more extreme parts, i.e., as SS increases. In the previous comparisons, we see that the SP does better than the LP in estimating λ. This occurs because, in the SP, we have a greater number of parts measured by all three operators, which more than offsets the value of leveraging.

In summary, by using a leveraged plan, we make substantial improvement in estimating the primary parameter of interest γ with no extra measurements. The only additional costs are the slightly more complex plan and the extra computational burden to find the MLE's and their standard errors.

4. LEVERAGED PLAN DESIGN

In this section, we present some general guidelines for choosing a leveraged plan (i.e., choosing values for b, k, and n) when the total number of measurements is $N = 60$ or 90 and there are $m = 3$ operators. We based our guidelines on designs that have the smallest asymptotic standard deviation of γ calculated using the Fisher information. To find the preferred plans with N and m fixed, for each point in the region $\gamma \in (0.01, 0.50)$ and $\lambda \in (0, 1)$, we calculate the asymptotic standard deviations for all possible designs. Note that when k is not a multiple of $2m$

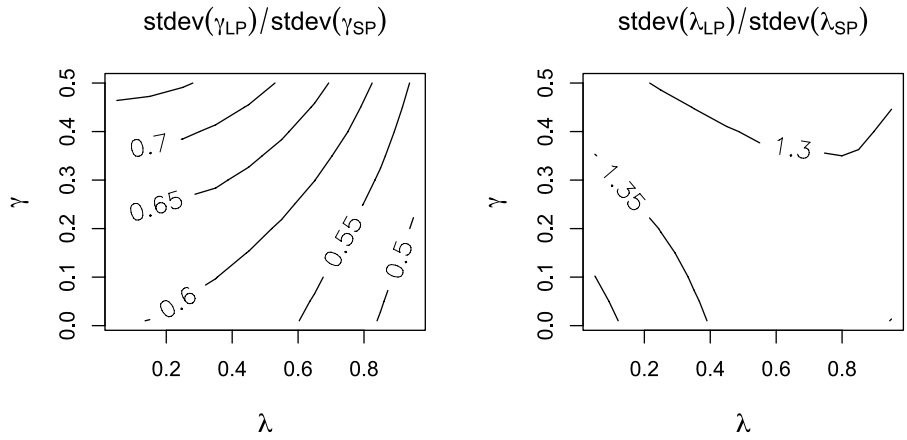

Figure 3. The ratios of standard deviations for the estimators of γ (left panel) and λ (right panel) from the leveraged plan ($N = 60$, $b = 11, k = 3, n = 3$) and standard plan ($N = 60, k = 10, n = 2$).

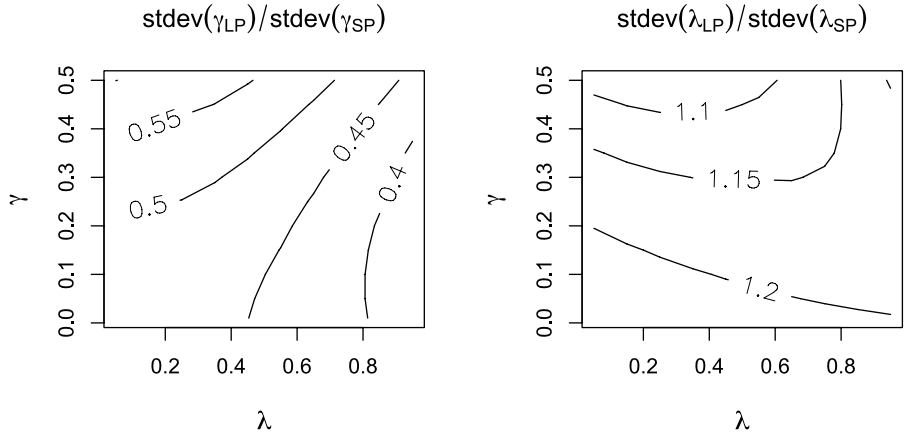

Figure 4. The ratios of standard deviations for the estimators of γ (left panel) and λ (right panel) from the leveraged plan ($N = 90$, $b = 18, k = 6, n = 2$) and standard plan ($N = 90, k = 10, n = 3$).

$(= 6)$, the set R of parts selected for Stage 2 cannot be balanced across the operators. In that case, to select the parts, we arbitrarily cycle through the operators taking the largest from the first, the smallest from the second and so on. We present the results of the search in Table 3.

The optimal designs in Table 3 change slowly with λ and γ. In practice, choosing an optimal design is not possible because of the dependency on the unknown parameters. We can suggest a criteria that summarizes the performance of a design over a range of reasonable values of γ and λ. However, any such criteria will be difficult to justify. In Table 3, we note in all cases the baseline size ($b \times m$) is around half the total number of measurements N and the number of repeated measurements n is two or three. To explore this observation further,

Table 3. The LP designs (b, k, n) with the smallest asymptotic standard deviation of γ given λ, γ when $m = 3$ and $N = 60$ (left) and $N = 90$ (right)

λ	$\gamma \in$	(b, k, n)	$\gamma \in$	(b, k, n)
0.1	(0.01, 0.50)	(10, 5, 2)	0.01	(16, 7, 2)
			(0.02, 0.26)	(15, 5, 3)
			(0.27, 0.50)	(16, 7, 2)
0.2	(0.01, 0.50)	(10, 5, 2)	(0.01, 0.09)	(16, 7, 2)
			(0.10, 0.29)	(15, 5, 3)
			(0.30, 0.50)	(16, 7, 2)
0.3	(0.01, 0.42)	(10, 5, 2)	(0.01, 0.19)	(16, 7, 2)
	(0.43, 0.50)	(12, 4, 2)	(0.20, 0.28)	(15, 5, 3)
			(0.29, 0.50)	(16, 7, 2)
0.4	(0.01, 0.22)	(10, 5, 2)	(0.01, 0.38)	(16, 7, 2)
	(0.23, 0.50)	(12, 4, 2)	(0.39, 0.50)	(18, 6, 2)
0.5	(0.01, 0.50)	(12, 4, 2)	(0.01, 0.26)	(16, 7, 2)
			(0.27, 0.50)	(18, 6, 2)
0.6	(0.01, 0.50)	(12, 4, 2)	(0.01, 0.06)	(16, 7, 2)
			(0.07, 0.50)	(18, 6, 2)
0.7	(0.01, 0.50)	(12, 4, 2)	(0.01, 0.40)	(18, 6, 2)
			(0.41, 0.50)	(20, 5, 2)
0.8	(0.01, 0.50)	(12, 4, 2)	(0.01, 0.04)	(18, 6, 2)
			(0.05, 0.50)	(20, 5, 2)
0.9	(0.01, 0.50)	(14, 3, 2)	(0.01, 0.50)	(22, 4, 2)

we consider the two leveraged plans ($b = 11, k = 3, n = 3$) and ($b = 18, k = 6, n = 2$) because they match designs featured in Table 3 and have k equal to an integer multiple of the number of operators. This choice of k allows an equal number of parts measured by each operator in the baseline to be represented in Stage 2. Figure 5 shows the asymptotic standard deviation of γ for these two plans versus the optimal designs given in Table 3 over the parameter space. For either plan, there is, at most, a 10% increase in the asymptotic standard deviation for γ from using the suggested LP design compared to the optimal design (which depends on the unknown parameters). Note also that the loss in efficiency is pronounced only when λ is close to 1.

We also investigated how these two plans behave when the operator effects are not equally spaced. Because we select the same number of parts for each operator from the baseline, we can show that the estimators from these plans do not depend on how the operator effects are distributed to make up σ_o^2.

We get similar results for other values of N and m. Thus, to summarize, we have the following guidelines for selecting a leveraged plan when only a few operators make regular use of the measurement system and the total number of measurements N available is roughly specified:

- Involve all m operators.
- In Stage 1, select $b \times m$ parts at random from the process where $b \times m$ is close to $N/2$. Each operator measures b different parts.
- In Stage 2, select an equal number of extreme parts from each operator based on the initial measurements to give a total of k parts. Since each of these parts will be measured two or three times by each operator, select k so that $2mk$ or $3mk$ is approximately $N/2$.

For example, suppose we have resources to make roughly $N = 50$ measurements and there are $m = 2$ operators. A near optimal plan is to select $b = 12$ so that the baseline has 24 parts. Then select the parts with the largest and smallest initial measured values from each operator so that $k = 4$. In Stage 2, each operator measures each of these parts $n = 3$ times for an overall total of 48 measurements. For estimating γ, this plan will be far better than the 12 part SP with each operator measuring each selected part twice.

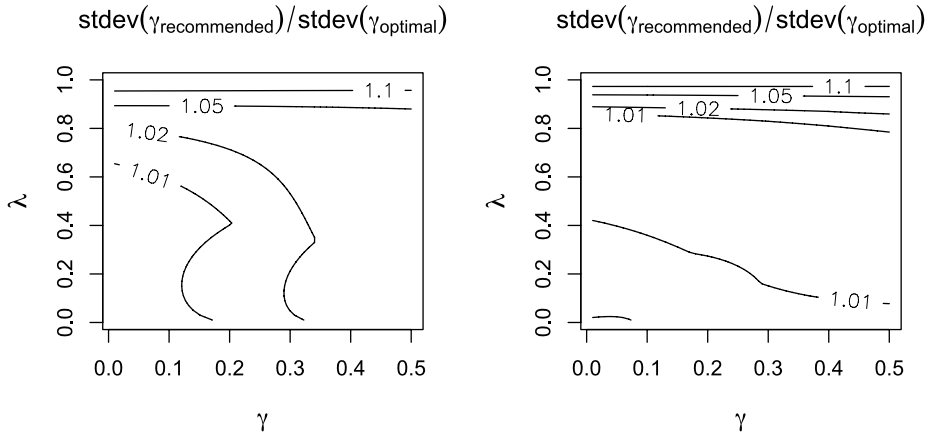

Figure 5. The standard deviation of the LP estimator from two designs $b = 11, k = 3, n = 3$ (left panel) and $b = 18, k = 6, n = 2$ (right panel) divided by the standard deviation of the optimal plan given $N = 60$ (left panel), 90 (right panel), ρ and $m = 3$.

We provide R code for the analysis and design for leveraged GR&R studies at *http://www.bisrg.uwaterloo.ca*.

5. DISCUSSION AND CONCLUSIONS

There many different ways to implement leveraging. When selecting parts for Stage 2, we suggest picking an equal number of extreme parts from each operator, and if k/m is even, choosing an equal number of large and small parts from each operator, i.e., if $m = 3$ and $k = 3$, we can pick the largest part from operators 1 and 3 and the smallest from operator 2. An alternative is to pick the most extreme parts across all operators, for example if $m = 3$ and $k = 3$, we can pick the three parts with the largest values of $|y_{ij0} - \bar{y}_{\cdot j0}|$. This allows for the possibility that all the parts for the Stage 2 can be measured by the same operator in the baseline. Choosing an even number of parts from each operator ensures balance, but the chosen parts are not likely to generate the largest value of SS, i.e., the most extreme measurements overall.

There are many alternative LP designs. In our suggested plan in the previous section, for $m = 3$ and $b = 10$, in the baseline stage we randomly selected 30 parts and allocate 10 to each operator. Another way to define the baseline is to randomly select 10 parts and have each operator measure each part once. This also gives a total of 30 measurements. Then, we might define the extreme parts for Stage 2 as having large or small averages (across operators). Qualitatively, it is clear that if we compare this version of an LP with the version suggested in Section 2, the information about λ will be increased at the expense of information about γ because we have fewer parts but more measurements from the different operators on the same part. We prefer the LP as suggested in Section 2 because our main goal is estimation of γ.

Typically, industrial measurement systems only have a few operators, but if we had a context where many operators use the system, we will prefer a model with random operator effects. One example, in a medical context, is measuring blood pressure. Many doctors measure their patient's blood pressure using the same process. In this case, with many operators, we can specify the baseline as b parts measured by b operators (one part per operator) for a total of b measurements. The baseline is defined in this way so that the sample average and standard deviation from the baseline are estimates of the overall average and standard deviation. Associated with each baseline measurement, we have both the operator and the part number. For Stage 2 where we select extreme parts, we have several options:

- Leverage by part: in Stage 2 repeatedly measure the chosen extreme parts with m new operators n times each.

  For example with $k = 2, m = 3$ and $n = 2$, we might pick the parts with the overall minimum and maximum initial measurements in the baseline sample. Then we recruit three new operators to measure these parts twice each.
- Leverage by operator: select the m operators associated with the extreme measurements in the baseline. Then, in Stage 2 these operators measure k new parts n times each.

  For example with $k = 2, m = 3$ and $n = 2$, we might pick the three operators associated with the minimum, maximum and second largest initial measurements in the baseline sample. Then, we obtain two new parts for these three operators to measure twice each.
- Leverage by part and operator: select the operator and part pairs associated with the extreme measurements in the baseline. Then in Stage 2, the chosen operators measure each of the selected parts n times each.

  For example, we might pick the parts and operators associated with the minimum and maximum initial measurements in the baseline sample so that $m = 2$, $k = 2$. Then, the two operators measure these two parts a further three times each.

We do not compare these possibilities further.

In summary, we give details on how to calculate the likelihood and Fisher information for the parameters of interest in a leveraged measurement system assessment plan. By tedious calculation, we show that as the level of leveraging increases, the asymptotic standard deviations of the MLE's for λ and γ decrease. We demonstrate the advantages of the LP over the standard plan using simulation. We show that the standard deviation of the LP estimator for γ, the primary parameter of interest, was significantly smaller than the corresponding estimator in an SP with the same total number of measurements. Conversely, the SP estimator for λ was shown to have a smaller

standard deviation than the LP. Good leveraged plans with a total of N measurements and m operators have bm roughly equal to $N/2$ where each operator measures b parts in the baseline. In Stage 2, select k extreme parts from the baseline so that k is an integer multiple of the number of operators and mkn is also roughly $N/2$ where n, the number of times each operator measures each selected part in Stage 2, is 2 or 3.

APPENDIX: THE ASYMPTOTIC VARIANCE–COVARIANCE MATRIX

In this appendix, we derive a detailed expression for the asymptotic variances of the MLE's for γ and λ by simplifying the result given in Equation (18). The object is to demonstrate that increasing $SS = \sum_{(i,j)\in R} E[Z_{ij0}^2]$, while holding $S_1, \ldots, S_m$ fixed, reduces the asymptotic variances of the estimates $\widehat{\gamma}$ and $\widehat{\lambda}$. That is, we demonstrate algebraically that leveraging is better than selecting random parts for Stage 2.

We start by showing that the Fisher matrix for $(\underline{\mu}, \sigma_{pg}^2, \rho)$ has a block matrix form

$$\mathbf{F} = \begin{pmatrix} \mathbf{M} & \mathbf{B} \\ \mathbf{B}^t & \mathbf{P} \end{pmatrix}, \quad \text{where } \mathbf{B} = (\,\mathbf{0}_m \quad \mathbf{v}\,).$$

Note that the only dependency on SS is through the [2, 2] element of the matrix $\mathbf{P}$ that is linear in SS. The vector $\mathbf{v}$ depends on $S_1, \ldots, S_m$ and is $\mathbf{0}$ for the balanced plan. For a general block matrix, its inverse is can be found in Rao (1973). Applying this result to $\mathbf{F}$ we obtain

$$\mathbf{F}^{-1} = \begin{bmatrix} \mathbf{M}^{-1} + \mathbf{K} & -\mathbf{M}^{-1}\mathbf{B}\mathbf{Q}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{B}^{\mathbf{t}}\mathbf{M}^{-1} & \mathbf{Q}^{-1} \end{bmatrix},$$

where

$$\mathbf{Q} = \mathbf{P} - \mathbf{B}^{\mathbf{t}}\mathbf{M}^{-1}\mathbf{B}. \tag{A.1}$$

Now the [2, 2] element of the matrix $\mathbf{Q}$ depends linearly on SS. The matrix $\mathbf{D}$ from Equation (17) can be written as

$$\mathbf{D} = \begin{pmatrix} \mathbf{I}_m & \mathbf{E} \\ \mathbf{0}_{m\times 2}^t & \mathbf{T} \end{pmatrix},$$

where

$$E = (\,\mathbf{w} \quad \mathbf{0}_m\,) \qquad \text{and} \qquad \mathbf{T} = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix}.$$

The inverse of $\mathbf{D}$ is

$$\mathbf{D}^{-1} = \begin{bmatrix} \mathbf{I} & -\mathbf{E}\mathbf{T}^{-1} \\ \mathbf{0}_{m\times 2}^t & \mathbf{T}^{-1} \end{bmatrix}.$$

A closed form expression for the asymptotic variance of $\widehat{\gamma}$ and $\widehat{\lambda}$ can be determined by setting

$$\mathbf{Q}^{-1} = \frac{1}{\det(\mathbf{Q})} \begin{pmatrix} q_{22} & -q_{12} \\ -q_{12} & q_{11} \end{pmatrix} \qquad \text{and} \qquad \alpha = \mathbf{v}^t\mathbf{M}^{-1}\mathbf{w},$$

and evaluating the product $(\mathbf{D}^t)^{-1}\mathbf{F}^{-1}\mathbf{D}^{-1}$. After some simplification, we obtain an expression for the asymptotic variance of $\widehat{\gamma}$ as

$$\frac{1}{\det(\mathbf{T})^2}\frac{t_{12}^2}{q_{11}}\left[\frac{q_{11}^2}{\det(\mathbf{Q})}\left(\frac{t_{11}}{t_{12}} + \frac{q_{12}}{q_{11}} - \alpha\right)^2 + 1 + q_{11}(\mathbf{w}^t\mathbf{M}^{-1}\mathbf{w})\right]. \tag{A.2}$$

If we increase SS, holding $S_1, \ldots, S_m$, n, m, b, and k fixed, then $\det(\mathbf{Q})$ increases and the asymptotic variance of $\widehat{\gamma}$ decreases. A similar calculation shows that the asymptotic variance of $\widehat{\lambda}$ also decreases as SS increases with the other design parameters held fixed.

SUPPLEMENTAL MATERIALS

Additional comparisons: Additional comparisons of leveraged to standard plan with 2 and 4 operators. (sup.pdf)

ACKNOWLEDGMENTS

The authors thank the editor, the associate editor, and the referees for many helpful suggestions to improve the article.

[Received March 2009. Revised April 2010.]

REFERENCES

Automotive Industry Action Group (2003), *Measurement Systems Analysis* (3rd ed.), Southfield, MI: Automotive Industry Action Group. [294,295]

Browne, R. P., MacKay, R. J., and Steiner, S. H. (2009a), "Improved Measurement System Assessment for Processes With 100% Inspection," *Journal of Quality and Technology*, 41, 376–388. [294]

——— (2009b), "Two-Stage Leveraged Measurement System Assessment," *Technometrics*, 51, 239–249. [294]

Burdick, R., Borror, C., and Montgomery, D. (2005), *Design and Analysis of Gauge R&R Studies: Making Decisions With Confidence Intervals in Random and Mixed ANOVA*. *ASA–SIAM Series on Statistics and Applied Probability*, Philadelphia, PA: ASA–SIAM. [294,295,299]

——— (2003), "A Review of Methods for Measurement Systems Capability Analysis," *Journal of Quality Technology*, 35, 342–354. [295]

Dillon, W., and Goldstein, M. (1984), *Multivariate Analysis Methods and Applications*, New York: Wiley. [296]

Dolezal, K. K., Burdick, R. K., and Birch, N. J. (1998), "Analysis of a Two-Factor R&R Study With Fixed Operators," *Journal of Quality Technology*, 30, 163–170. [294,295]

Montgomery, D. C., and Runger, G. (1993a), "Gauge Capabilty Analysis and Designed Experiments. Part I: Basic Methods," *Quality Engineering*, 6, 115–135. [294]

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications* (2nd ed.), New York: Wiley. [302]

Wheeler, D. J., and Lyday, R. (1984), *Evaluating the Measurement Process*, Knoxville, TN: SPC Press. [294]