Two-Stage Leveraged Measurement System Assessment

Ryan P. Browne, R. Jock MacKay, and Stefan H. Steiner

Department of Statistics and Actuarial Sciences
Business and Industrial Statistics Research Group
University of Waterloo
Waterloo, N2L 3G1 Canada
(*rpbrowne@uwaterloo.ca*; *rjmackay@uwaterloo.ca*; *shsteine@uwaterloo.ca*)

To assess measurement system variation, we propose an alternative to the standard plan that uses a random sample of parts to repeatedly measure. The new plan, called Leveraged Measurement System Assessment is conducted in two stages. In the first stage, called the baseline, a number of parts are measured once. In the second stage, we select a few extreme parts (based on their initial measurement in the baseline) and remeasure each of them a number of times. We demonstrate the advantage of the leveraged over the standard plan by comparing the bias and standard deviation of estimators of the intraclass correlation coefficient. We also present a method to determine sample size when planning a Leveraged Measurement System Assessment.

KEY WORDS: Intraclass correlation; Leverage; Measurement system assessment.

1. INTRODUCTION

Good measurement systems are critical in a manufacturing environment to allow control and to support process improvement and decision making. More generally, measurement systems are needed in any scientific inquiry or quantitative decision-making process. As a result, it is important to have efficient methods to assess the properties of measurement systems.

To assess the variability of a measurement system, we repeatedly measure a number of parts. We adopt the common random effects model

$$Y_{ij} = P_i + E_{ij}, \qquad i = 1, 2, \ldots, k \text{ and } j = 1, 2, \ldots, n, \quad (1)$$

where P_i is a random variable representing the true dimension of part i, E_{ij} is a random variable representing the error on the jth measurement on the ith part, n is the number of repeated measurements on each part, and k is the number of parts. We assume that the part effects P_i are independent and identically distributed normal random variables with mean μ and variance σ_p^2, the measurements errors E_{ij} are independent and identically distributed normal random variables with mean zero and variance σ_m^2, and P_i and E_{ij} are mutually independent. The variance of Y_{ij}, called the total variation, is $\sigma_t^2 = \sigma_p^2 + \sigma_m^2$. By adopting model (1), we assume that μ, σ_p, and σ_m are constant over the time needed to conduct the measurement assessment investigation.

With model (1), we also assume the studied measurement system has no operator effects. This occurs with automated measurement systems. For example, in one application piston diameters are measured by an inline gauge with automated part handling. Operators play no part in the operation of the gauge. Using manufacturing jargon, in our context, σ_m captures measurement repeatability but not reproducibility.

To quantify the contribution of the measurement system to the total variation, we use the intraclass correlation coefficient, ρ, defined as the ratio of the process variation over the total variation, that is, $\rho = \sigma_p^2 / \sigma_t^2$. Note that $0 \le \rho \le 1$ and the larger the value of ρ, the smaller is the contribution of the measurement system to the overall variation.

The intraclass coefficient is equivalent to gauge repeatability $\gamma = \sigma_m / \sigma_t$ (Automotive Industry Action Group 2002) where smaller values are better. The commonly used cut-off values to determine if a measurement system is acceptable are $\gamma = 0.10$ and $\gamma = 0.30$. These values correspond to ρ equal to 0.99 and 0.91, respectively.

A standard measurement assessment plan (SP) is to sample k parts selected at random from the process and measure them n times each for a total of $N = nk$ measurements (Automotive Industry Action Group 2002). The SP typically uses two or three operators, each of whom measure 10 parts two or three times for a total of 40 to 90 measurements. In the context considered here, there is no need to use different operators. Throughout the paper, we use a standard plan with 10 parts and 6 repeated measurements on each part as a basis of comparison. This plan produces the data y_{ij}, where $i = 1, \ldots, k$ and $j = 1, \ldots, n$. Following Donner and Eliasziw (1987), we estimate ρ by the sample intraclass correlation coefficient

$$\frac{MSA - MSW}{MSA - (n-1)MSW}, \quad (2)$$

where MSA and MSW are

$$MSA = \sum_{i=1}^{k} n(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 / (k-1),$$

$$MSW = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i\cdot})^2 / [k(n-1)]$$

and

$$\overline{y}_{i\cdot} = \sum_{j=1}^{n} y_{ij}/n, \qquad \overline{y}_{\cdot\cdot} = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}/nk.$$

In this article, we introduce a new leveraged plan (LP) where we deliberately select extreme parts to remeasure from an initial baseline sample. We demonstrate the advantages of the LP over the standard plan (SP) by comparing the standard deviation and bias of a number of estimators for ρ.

The paper is structured as follows. In the next section, we describe the LP in more detail. In Section 3, we present four estimates of ρ for the LP, give a numerical example, and compare the properties of the four estimators for a particular LP. In Section 4, we compare the standard and leveraged plans using the bias and standard deviation of the derived estimators. In Section 5, we compare different designs for LPs when the total sample size is fixed. Based on empirical evidence, we recommend a specific plan for any total sample size. In Section 6, we look at methods for assessing the assumptions required for a LP to be effective and also briefly look at the issue of robustness. Finally, in Section 7, we provide some discussion and possible extensions to the LP discussed here.

2. THE LEVERAGED MEASUREMENT ASSESSMENT PLAN

We conduct a leveraged measurement system assessment in two stages:

Stage 1: Sample b parts at random from the process to obtain a baseline. We denote the observed values $\{y_{10}, y_{20}, \ldots, y_{b0}\}$ and the baseline average and sample variance by $\overline{y}_b = \frac{1}{b}\sum_{i=1}^{b} y_{i0}$ and $s_b^2 = \frac{1}{b-1}\sum_{i=1}^{b}(y_{i0} - \overline{y}_b)^2$.

Stage 2: From the baseline sample, select k parts using the observed measured values. In particular, to improve the estimation of ρ, sample k parts so that:

- the initial measurements are extreme relative to the baseline average, and
- the average of the initial measurements of the selected parts is close to the baseline average.

We denote the k selected parts using the set R. These k parts are then repeatedly measured n times each to give the additional data $\{y_{ij}, i \in R \text{ and } j = 1, \ldots, n\}$. The total number of measurements in the LP is $N = b + nk$.

For example, for a LP with $k = 2$, we might pick the parts with the minimum and maximum initial measurement in the baseline sample. In Stage 1, 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. This recommendation matches the requirements for the SP (Automotive Industry Action Group 2002).

We use the term leveraged plan because of the reuse of units with extreme values. In a more general context, the term leverage is sometimes used by the proponents of the problem solving system initially proposed by Dorian Shainin. Units with relatively large and small values of the response are compared to identify the major causes of the variation. See Steiner, MacKay, and Ramberg (2008) for a fuller description.

3. ESTIMATES FOR THE LEVERAGED PLAN

We present and compare four estimates of ρ. The first, here called the ANOVA estimate, is straightforward. It is based on (2) using the variation within the repeated measurements to estimate σ_m and the variation from the baseline to estimate σ_t. The ANOVA estimate does not benefit from leveraging; in fact, we lose one degree of freedom per part because we do not include the baseline measurement in the calculation of *MSW* in (2). The second is the maximum likelihood estimate (MLE) that must be determined by numerical approximation. The third estimate uses a regression approach that exploits the fact that the conditional mean of the repeated measurements, given the initial measurement, depends linearly on ρ. Finally, the fourth estimate is a linear combination of the ANOVA and regression estimates.

Note that the three alternatives to the MLE use the baseline data only to estimate μ and σ_t^2 and then use the repeated measurements, conditional on the baseline data, to estimate ρ.

3.1 ANOVA Estimate

We use the variation within the repeated measurements to get an ANOVA-like estimate of ρ. For each part i in R, the variation within the repeated measurements $\sum_{j=1}^{n}(Y_{ij} - \overline{Y}_{i\cdot})^2$ is independent of Y_{i0} and

$$MSW = \frac{\sum_{i \in R}\sum_{j=1}^{n}(y_{ij} - \overline{y}_{i\cdot})^2}{k(n-1)} \qquad (3)$$

is an estimate of σ_m^2. Note that the average $\overline{y}_{i\cdot}$ in the above expression does not include the baseline measurements. Since the baseline variance s_b^2 is an estimate of σ_t^2 and $\rho = \sigma_p^2/\sigma_t^2$, by rearrangement, we obtain the ANOVA estimate

$$\widehat{\rho}_a = 1 - \frac{MSW}{s_b^2}. \qquad (4)$$

Transforming the corresponding estimator, we see that $(1 - \widehat{\rho}_a)/(1 - \rho)$ has an F-distribution with $k(n-1)$ and $b - 1$ degrees of freedom and so the distribution of the ANOVA estimator depends only on ρ and not the other unknown parameters μ and σ_t. We have

$$E(\widehat{\rho}_a) = 1 - (1 - \rho)\frac{b-1}{b-3} = \rho\left(\frac{b-1}{b-3}\right) - \frac{2}{b-3}, \qquad (5)$$

$$\sigma_a^2 = \mathrm{Var}(\widehat{\rho}_a)$$
$$= (1 - \rho)^2\frac{2(b-1)^2[k(n-1) + (b-1) - 2]}{k(n-1)[(b-1) - 2]^2[(b-1) - 4]}. \qquad (6)$$

Note that the ANOVA estimator does not require the selected parts in R to be representative of the process. It does, however, require that the measurement errors be representative which might be untrue for atypical parts; see Section 6.

3.2 Maximum Likelihood Estimate

For a single part i selected at random, the joint distribution of the baseline measurement and the n repeated measurements is

$$\begin{pmatrix} Y_{i0} \\ Y_{i1} \\ \vdots \\ Y_{in} \end{pmatrix} \sim N\left(\mu \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \sigma_t^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \\ \vdots & & \ddots & \vdots \\ \rho & & \cdots & 1 \end{bmatrix} \right). \quad (7)$$

The distribution of the repeated measurements $\{Y_{i1}, \ldots, Y_{in}\}$ conditional on the baseline measurement $Y_{i0} = y_{i0}$ is

$$\begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in} \end{pmatrix} \Bigg| Y_{i0} = y_{i0} \sim N\left(\mu + \rho(y_{i0} - \mu)\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \right.$$

$$\left. \Sigma = \sigma_t^2 \begin{bmatrix} 1 - \rho^2 & & \rho(1-\rho) \\ & \ddots & \\ \rho(1-\rho) & & 1 - \rho^2 \end{bmatrix} \right).$$
$$(8)$$

The covariance matrix Σ has a special form which allows us to obtain the following well-known properties (Dillon and Goldstein 1984):

$$\Sigma^{-1} = \frac{1}{\sigma_t^2(1-\rho)(1+n\rho))}$$
$$\times \begin{bmatrix} 1 + \rho(n-1) & & -\rho \\ & \ddots & \\ -\rho & & 1 + \rho(n-1) \end{bmatrix},$$

$$|\Sigma| = \sigma_t^{2n}(1-\rho)^n(1+n\rho).$$

Using the properties of Σ, we can write the conditional likelihood (conditional on y_0) for the repeated measurements on a single part. The measurements for one part are independent of the measurements from another part; thus the conditional likelihood for k parts, each with n measurements, is the product of the individual likelihoods. The conditional log-likelihood for n repeated measurements on k parts is

$$l_r(\mu, \sigma_t^2, \rho | y_{10}, \ldots, y_{b0})$$
$$= -\frac{nk}{2}\log\sigma_t^2 - \frac{nk}{2}\log(1-\rho) - \frac{k}{2}\log(1+n\rho)$$
$$- \frac{1}{2}\frac{1}{\sigma_t^2(1-\rho)(1+n\rho)}$$
$$\times \left\{ (1+n\rho)k(n-1)MSW \right.$$
$$\left. + n\sum_{i\in R}[\bar{y}_{i\cdot} - \mu - \rho(y_{i0} - \mu)]^2 \right\},$$

where MSW is defined, as in the ANOVA estimate, by (3) and the part average $\bar{y}_{i\cdot}$ omits the baseline measurement y_{i0}. Assuming the b parts in the baseline sample are selected at random from the process, the marginal log-likelihood of the baseline is

$$l_0(\mu, \sigma_t^2) = -\frac{b}{2}\log\sigma_t^2 - \frac{1}{2\sigma_t^2}\{(b-1)s_b^2 + b(\bar{y}_b - \mu)^2\}.$$

Thus, the (unconditional) log-likelihood for the LP is

$$l(\mu, \sigma_t^2, \rho) = l_0(\mu, \sigma_t^2) + l_r(\mu, \sigma_t^2, \rho | y_{10}, \ldots, y_{b0}). \quad (9)$$

This likelihood is appropriate regardless of how we select the parts to be remeasured in Stage 2 of the LP; see Appendix A. To get the MLEs of μ, σ_t^2 and ρ, we maximize (9). Solutions can be found numerically.

The asymptotic variance–covariance matrix of the maximum likelihood estimators is the inverse of the Fisher information matrix, given by

$$J(\mu, \sigma_t^2, \rho) = \begin{pmatrix} \frac{(1-\rho)nk+b(n\rho+1)}{\sigma_t^2(n\rho+1)} & 0 \\ 0 & \frac{1}{2}\frac{b+nk}{\sigma_t^4} \\ \frac{nE[SC]}{\sigma_t(n\rho+1)} & -\frac{1}{2}\frac{nk\rho(n+1)}{\sigma_t^2(n\rho+1)(1-\rho)} \\ & \frac{nE[SC]}{\sigma_t(n\rho+1)} \\ & -\frac{1}{2}\frac{nk\rho(n+1)}{\sigma_t^2(n\rho+1)(1-\rho)} \\ & E[-\frac{\partial^2}{\partial\rho^2}l(\mu, \sigma_t^2, \rho)] \end{pmatrix}, \quad (10)$$

where

$$E\left[-\frac{\partial^2}{\partial\rho^2}l(\mu, \sigma_t^2, \rho)\right] = \frac{1}{2}\frac{kn^2}{(1+n\rho)^2} + \frac{kn\rho(n+1)}{(1+n\rho)(1-\rho)^2}$$
$$- \frac{1}{2}\frac{kn}{(1-\rho)^2} + \frac{nE[SSC]}{(1-\rho)(1+n\rho)},$$

$$SSC = \sum_{i\in R}\left[\frac{Y_{i0} - \mu}{\sigma_t}\right]^2, \quad \text{and} \quad SC = \sum_{i\in R}\left[\frac{Y_{i0} - \mu}{\sigma_t}\right].$$

We show in Appendix B that if a sampling plan is chosen such that $E[SC] = 0$ and $E[SSC]$ is large, then the asymptotic variance of the MLE for ρ is reduced. This observation is the reason for the Stage 2 recommendations in the definition of the LP in Section 2. A plan with both these properties is to choose an equal number of parts with extreme baseline measurements on either side of the baseline average.

When using maximum likelihood, standard errors for the estimates can be obtained from the inverted information matrix (see Appendix B) with the parameters replaced by their estimates.

3.3 Regression Estimate

We now derive an alternative to maximum likelihood that has a closed form using a regression model. From (8), the distribution of the average of the repeated measurements on a single part, given the baseline measurement is

$$\bar{Y}_{i\cdot}|(Y_{i0} = y_{i0}) \sim N\left(\mu + \rho(y_{i0} - \mu), \sigma_t^2(1-\rho)\left(\rho + \frac{1}{n}\right)\right). \quad (11)$$

The averages of the repeated measurements on different parts are mutually independent. Since in (11), the mean depends on ρ linearly and the variance is the same for each part, we can use regression to estimate ρ. The conditional mean of $\bar{Y}_{i\cdot}$ also depends on μ but we use the baseline average $\bar{y}_b$ to estimate this unknown.

The regression estimate of ρ (Montgomery, Peck, and Vining 2001) is

$$\widehat{\rho}_r = \frac{\sum_{i\in R}(\bar{y}_{i\cdot} - \bar{y}_b)(y_{i0} - \bar{y}_b)}{\sum_{i\in R}(y_{i0} - \bar{y}_b)^2}. \quad (12)$$

If we standardize each quantity in (12), we see that the marginal distribution of $\hat{\rho}_r$ depends only on ρ. The distribution of $\hat{\rho}_r$, conditional on the baseline sample, is normal with mean

$$E[\hat{\rho}_r | y_{10}, \ldots, y_{b0}] = \rho + \left[(\mu - \bar{y}_b) \frac{\widehat{SC}}{\widehat{SSC}} \right] (1 + \rho) \qquad (13)$$

and variance

$$\text{Var}[\hat{\rho}_r | y_{10}, \ldots, y_{b0}] = \frac{\sigma_t^2 (1 - \rho)(1/n + \rho)}{s_b^2 \widehat{SSC}},$$

where

$$\widehat{SC} = \sum_{i \in R} \left[\frac{y_{i0} - \bar{y}_b}{s_b} \right] \quad \text{and} \quad \widehat{SSC} = \sum_{i \in R} \left[\frac{y_{i0} - \bar{y}_b}{s_b} \right]^2 \qquad (14)$$

are the baseline estimates of SC and SSC as defined in (10).

The estimator $\hat{\rho}_r$ has a small bias (conditionally) because we recommend choosing parts in Stage 2 so that $\widehat{SC} \approx 0$ and $\widehat{SSC}$ is large. We also expect that $\bar{y}_b$ will be close to μ since the baseline sample is selected at random from the process. Unconditionally, $\hat{\rho}_r$ is unbiased because $\overline{Y}_b$ is independent of the random variables corresponding to $\widehat{SC}$ and $\widehat{SSC}$.

The unconditional variance of $\hat{\rho}_r$ is

$$\sigma_r^2 = \text{Var}(\hat{\rho}_r)$$

$$\approx (1 - \rho) \left(\rho + \frac{1}{n} \right) E \left[\frac{\sigma_t^2}{\sum_{i \in R} (Y_{i0} - \overline{Y}_b)^2} \right] \qquad (15)$$

because $\hat{\rho}_r$ is approximately conditionally unbiased. We estimate $E[\sigma_t^2/(\sum_{i \in R}(Y_{i0} - \overline{Y}_b)^2)]$ from the baseline observations with the inverse of $\widehat{SSC}$ as given by (14). As with the MLE, choosing parts to remeasure with extreme baseline measurements with average close to the $\overline{y}_b$ reduces the variance of the regression estimator.

Note that the regression estimator uses the average but not the variability of the repeated measurements to estimate ρ unlike the next estimator.

3.4 Combined Estimate

An estimate which has a closed form and turns out to have properties similar to the MLE is a combination of the two estimators $\hat{\rho}_a$ and $\hat{\rho}_r$ as described in Sections 3.1 and 3.3, respectively. We can show that, given the baseline data, these two estimators are conditionally independent. Furthermore, as seen in (13), the estimator $\hat{\rho}_r$ is nearly unbiased. Hence the marginal covariance of $\hat{\rho}_r$ and $\hat{\rho}_a$ is close to 0. In what follows, we ignore this covariance. The combined estimator is a linear combination of $\hat{\rho}_r$ and $\hat{\rho}_a$ with weights selected to minimize the variance.

In general, if we have two unbiased independent estimators $\hat{\rho}_r$ and $\hat{\rho}_a$ with known variances σ_a^2 and σ_r^2, the minimum variance linear combination is

$$w\hat{\rho}_r + (1 - w)\hat{\rho}_a = \frac{\sigma_a^2}{\sigma_r^2 + \sigma_a^2} \hat{\rho}_r + \frac{\sigma_r^2}{\sigma_r^2 + \sigma_a^2} \hat{\rho}_a. \qquad (16)$$

This combined estimator is approximately unbiased because it is a weighted sum of two approximately unbiased estimators, $\hat{\rho}_a$ and $\hat{\rho}_r$. An estimating function can be created from (16) by subtracting its expectation. Multiplying by $\sigma_r^2 + \sigma_a^2$, we get

$$\Psi_c(\rho) = \sigma_a^2 \hat{\rho}_r + \sigma_r^2 \hat{\rho}_a - (\sigma_a^2 + \sigma_r^2)\rho. \qquad (17)$$

To find $\hat{\rho}_c$, we replace $\hat{\rho}_r$ and $\hat{\rho}_a$ by the corresponding estimates, substitute the quantities in (6) for σ_a^2 and (15) for σ_r^2 and set the corresponding function to zero. We obtain the combined estimate $\hat{\rho}_c$, as a root of the quadratic equation (18).

$$\left(v_F - E \left[\frac{1}{SSC} \right] \right) \rho_c^2$$

$$+ \left(E \left[\frac{1}{SSC} \right] \left[\hat{\rho}_a - \frac{1}{n} \right] - v_F[1 + \hat{\rho}_r] \right) \rho_c$$

$$+ \left(v_F \hat{\rho}_r + E \left[\frac{1}{SSC} \right] \frac{\hat{\rho}_a}{n} \right) = 0, \qquad (18)$$

where $v_F = \sigma_a^2/(1 - \rho)^2$. As with the regression estimate, we estimate $E[\frac{1}{SSC}]$ from the baseline data with the inverse of $\widehat{SSC}$ as given by (14).

In this case, the appropriate estimate is the smaller root because the larger root is greater than one. Note that $\hat{\rho}_c$ is not just a simple weighted average of the two previous estimates because the variances σ_r^2 and σ_a^2 depend on ρ.

From Jorgensen and Knudsen (2004), the asymptotic variance of the combined estimator is approximately

$$\text{Var}(\hat{\rho}_c) \approx \frac{\text{Var}[\Psi_c(\rho)]}{E[\frac{\partial}{\partial \rho} \Psi_c(\rho)]^2} = \frac{\sigma_a^2 \sigma_r^2}{(\sigma_a^2 + \sigma_r^2)}. \qquad (19)$$

The asymptotic variance covariance matrix (see Jorgensen and Knudsen 2004) of $\hat{\mu}, \hat{\sigma}_t^2$, as estimated from the baseline, and $\hat{\rho}_c$, as given by solving (18), is

$$\text{Var} \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}_t^2 \\ \hat{\rho}_c \end{pmatrix} \approx \begin{pmatrix} \frac{\sigma_t^2}{b} & 0 & 0 \\ 0 & \frac{2\sigma_t^4}{b-1} & \frac{2\sigma_t^2(1-\rho)}{b-3} \frac{\sigma_r^2}{\sigma_r^2 + \sigma_a^2} \\ 0 & \frac{2\sigma_t^2(1-\rho)}{b-3} \frac{\sigma_r^2}{\sigma_r^2 + \sigma_a^2} & \frac{\sigma_a^2 \sigma_r^2}{(\sigma_a^2 + \sigma_r^2)} \end{pmatrix}. \qquad (20)$$

The variance of the combined estimator depends on ρ through σ_a^2 and σ_r^2.

3.5 Numerical Example of Various Estimates for ρ From an LP

Steiner and Mackay (2005) present an example of a leveraged measurement assessment study. Although they calculate only the ANOVA estimator, we can apply all four methods of estimation for illustration. In the example, three parts, a large, small, and medium-sized part, were selected from the baseline study to be remeasured. To more closely match the suggestions in this paper to select an equal number of extreme parts on each side of the baseline average, we proceed assuming only one large and one small part were selected. A baseline of 100 parts was randomly selected from the process. The baseline data, given as a difference from nominal, are shown in Table 1.

The baseline average $\bar{y}_b$ is 0.540 and baseline variance is s_b^2 is 25.865. The parts chosen in Stage 2 were parts 50 and 70 (i.e., $R = \{50, 70\}$), with baseline measurements 12.8 and -12.2, respectively. These two parts were measured an additional 18 times each. Note that the average of the two baseline measurements for the selected parts is very close to the baseline average. Part 21 with baseline value -12.8 could have been selected instead of part 70 but for reasons unknown to us, this was not done.

Table 1. Baseline data of 100 camshaft journal diameters

5.3	0.0	−4.1	−6.4	−5.7	7.1	−0.5	−1.7	−2.7	2.1
0.9	−1.5	−5.4	3.3	6.0	2.4	−1.2	3.4	−2.9	−6.4
−12.8	−7.3	1.5	1.9	5.6	−5.2	2.4	0.9	−2.5	−0.8
4.6	4.1	−7.8	10.3	0.0	−0.9	−3.3	5.7	8.2	1.5
−5.3	4.2	4.6	10.5	−3.4	0.5	1.4	9.1	−1.1	**12.8**
−2.7	−3.2	4.4	1.0	1.2	−4.0	−1.6	−2.5	−6.9	1.2
−2.2	−0.6	−5.4	−6.0	−1.1	0.1	−3.5	2.5	1.4	**−12.2**
−1.5	−6.0	9.7	5.2	10.4	2.2	9.2	3.6	1.8	1.7
−2.0	−0.8	−4.1	−4.5	4.2	7.8	−3.2	1.9	−0.4	0.5
4.3	2.3	6.1	5.0	4.6	8.4	6.1	−7.1	4.7	−7.4

The individual measurements for each remeasured part are shown in Table 2 with the average and standard deviation for the repeated measurements within each part. We see that the measurement system is easily able to distinguish the two parts and that the measurement variation for the two parts is roughly the same. There is no evidence to contradict the model assumptions in this example. It is interesting to note (as pointed out by an observant referee) that all of the remeasured values for part 70 are larger than the baseline value. This suggests that there was a large measurement error in the baseline value of this part. This phenomenon is more likely to happen if the measurement variation is relatively large compared to the process variation.

To check the normality assumptions, we constructed QQ-plots (not shown here) of the baseline data and the residuals of the repeated measurements excluding the baseline measurements. There is no evidence agianst the normality assumptions. Also, the sample variances s_{50}^2 and s_{70}^2 are not significantly different.

Using (3), the ANOVA estimate is

$$\widehat{\rho}_a = 1 - \frac{MSW}{s_b^2} = 1 - \frac{(s_{50}^2 + s_{70}^2)/2}{s_b^2}$$

$$= 1 - \frac{(0.40997 + 0.68029)/2}{25.865} = 0.97892.$$

The estimates of SC and SSC using the two selected parts and the baseline summary statistics are

$$\widehat{SC} = -2.51 + 2.41 = -0.10 \qquad \text{and}$$

$$\widehat{SSC} = 6.275 + 5.811 = 12.086.$$

The maximum likelihood estimates for (μ, σ_t^2, ρ) are (0.551, 25.392, 0.97809).

Using (12), the regression estimate is

$$\widehat{\rho}_r = \frac{\sum_{i \in R}(\bar{y}_{i\cdot} - \bar{y}_b)(y_{i0} - \bar{y}_b)}{\sum_{i \in R}(y_{i0} - \bar{y}_b)^2}$$

$$= \frac{145.8 + 148.9}{162.3 + 150.3} = \frac{294.7}{312.6} = 0.94267.$$

We need v_F and the baseline estimate of SSC to determine the coefficients of the quadratic equation (18) used for the combined estimator. Using $v_F = \text{Var}(F_{34,99}) = 0.0845$, the combined estimate of ρ is the smaller root of the quadratic equation

$$0.001755011\rho_c^2 - 0.0877455\rho_c + 0.08414984 = 0.$$

The two roots of this equation are 0.97816 and 49.019 so we have $\widehat{\rho}_c = 0.97816$. Table 3 summarizes the four estimates and their corresponding standard errors. Since ρ appears to be large for this measurement system, there is little difference in the estimates and their standard errors with the exception of the regression estimate which has much higher standard error.

To create confidence intervals using an estimator with a skewed distribution, it is common to work on a transformed scale. Fisher's z transformation performs well here. We let

$$\theta = \frac{1}{2} \log \frac{1+\rho}{1-\rho}, \qquad \frac{\partial \theta}{\partial \rho} = \frac{1}{1-\rho^2},$$

$$\hat{\theta} = \frac{1}{2} \log \frac{1+\hat{\rho}}{1-\hat{\rho}} \tag{21}$$

and then

$$\text{Var}(\hat{\theta}) \approx \text{Var}(\hat{\rho}_c)\left[\frac{\partial \theta}{\partial \rho}\right]^2. \tag{22}$$

To create a confidence interval for ρ, we first create a confidence interval for θ and then transform the limits. We illustrate the calculations using the combined estimate. The transformed

Table 2. Example of a Stage 2 sample with 2 extreme parts repeatedly measured 18 times each

Part 50 $y_{500} = 12.8$						Part 70 $y_{700} = -12.2$					
10.9	13.2	12.8	12.6	12.7	14.1	−10.3	−11.1	−10.0	−12.2	−11.0	−11.1
12.9	13.1	12.0	13.3	12.6	13.4	−10.9	−10.0	−10.6	−11.4	−11.5	−11.1
12.0	12.9	11.7	11.8	12.2	14.1	−11.4	−10.7	−10.3	−11.4	−9.8	−11.5
		$\bar{y}_{50\cdot} = 12.7$ $s_{50}^2 = 0.68029$						$\bar{y}_{70\cdot} = -10.9$ $s_{70}^2 = 0.40997$			

Table 3. Estimates of ρ for the camshaft journal diameters example

	Estimate	Standard error
$\widehat{\rho}_a$	0.97892	0.00613
$\widehat{\rho}_r$	0.94267	0.06881
$\widehat{\rho}_c$	0.97816	0.00628
$\widehat{\rho_{MLE}}$	0.97809	0.00597

estimate, using (21) is $\frac{1}{2}\log(\frac{1+0.97816}{1-0.97816}) = 2.2531$. The standard error of this estimate is given in (22), using the standard error of the combined estimate (found in Table 3), is $\frac{0.00628}{(1-0.97816^2)} = 0.14535$. A 95% confidence interval for θ is $2.25 \pm 1.96(0.145) = (1.968, 2.538)$ and the approximate 95% confidence interval in terms of ρ is $(0.962, 0.988)$.

3.6   Comparison of Estimates of ρ in the LP

We consider a LP with $b = 30$, $k = 6$, and $n = 5$ because this is the plan that we will recommend in Section 5 when the total number of measurements is 60. For Stage 2, we chose the six parts which have the three largest and three smallest measurements from the baseline of 30 parts. For the simulation, we did not try to match the average of the selected parts with the baseline average. Figure 1 shows the bias and standard deviation for the ANOVA estimator ($\hat{\rho}_a$), the MLE, the regression estimator ($\hat{\rho}_r$), and the combined estimator ($\hat{\rho}_c$). The results of the simulation are based on 10,000 samples for each of 23 values of ρ spread over the interval (0.01, 0.99) with higher density where the bias and standard deviation are changing rapidly. We used the same set of values for ρ in all simulations.

The regression and ANOVA estimators are efficient, relative to the MLE, for different values of ρ which explains why the combined and the MLE estimators perform similarly when $\rho \geq 0.3$. Since the typical situation for a measurement system has ρ larger than 0.5, we can use the combined estimate without loss of efficiency. We see similar results for leveraged plans with other values of b, k, and n.
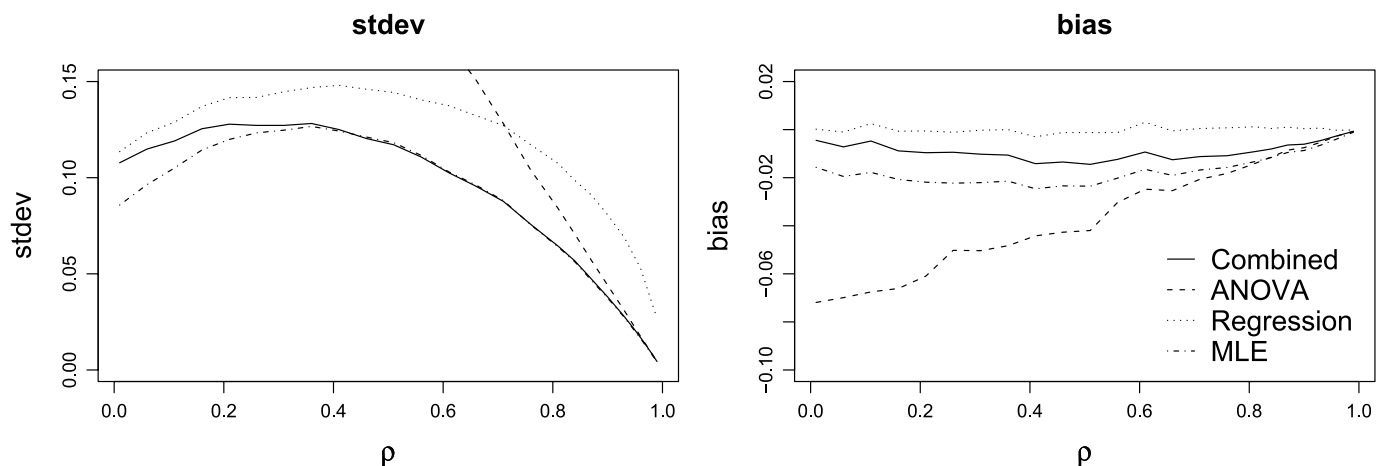
4.   LEVERAGED VERSUS STANDARD PLANS

To demonstrate the value of leveraging, we again resort to simulation. We compare the following two plans, each with a total of 60 measurements:

- SP with $k = 10$ and $n = 6$ (a commonly used plan in practice)
- LP with $b = 30$, $k = 6$, and $n = 5$ (as recommended in Section 5).

We use maximum likelihood estimation for both plans to make the comparisons fair. We quantify the difference between the plans using bias and standard deviation calculated from 10,000 simulations at 23 values of ρ spread over the interval (0.01, 0.99). We see from Figure 2 that the LP is substantially better than the SP with smaller standard deviation for all values of ρ and less bias when $\rho \geq 0.4$.

We can also compare the LP and SP by looking at the total number of measurements required to give a desired precision in the estimation of ρ. Figure 3 shows the total sample size required for a LP to have the same precision (standard deviation) as the SP ($k = 10$, $n = 6$) for different values of ρ. In Figure 3, the selected LP corresponds to the suggested plan from Section 5. For example, at $\rho = 0.91$, the SP has a standard deviation of 0.060 (see Figure 2). The LP with the same standard deviation for estimating ρ has a total sample size of 34, where $k = 3$, $n = 5$, and $b = 19$. We see similar results for other choices of n and k in the SP.

We conclude from the comparisons that the two-stage leveraged plan provides a substantial benefit over the standard plan.

5.   LEVERAGED PLAN DESIGN

In this section, we show how to design a LP (i.e., choose values for b, k, and n) when the total number of measurements is N and the precision desired for the estimate of ρ using the combined estimator are specified. Here k corresponds to the parts with the most extreme baseline values relative to the baseline average. As with most sample size calculations, we must also specify a value of ρ to select the plan. We consider two spe-
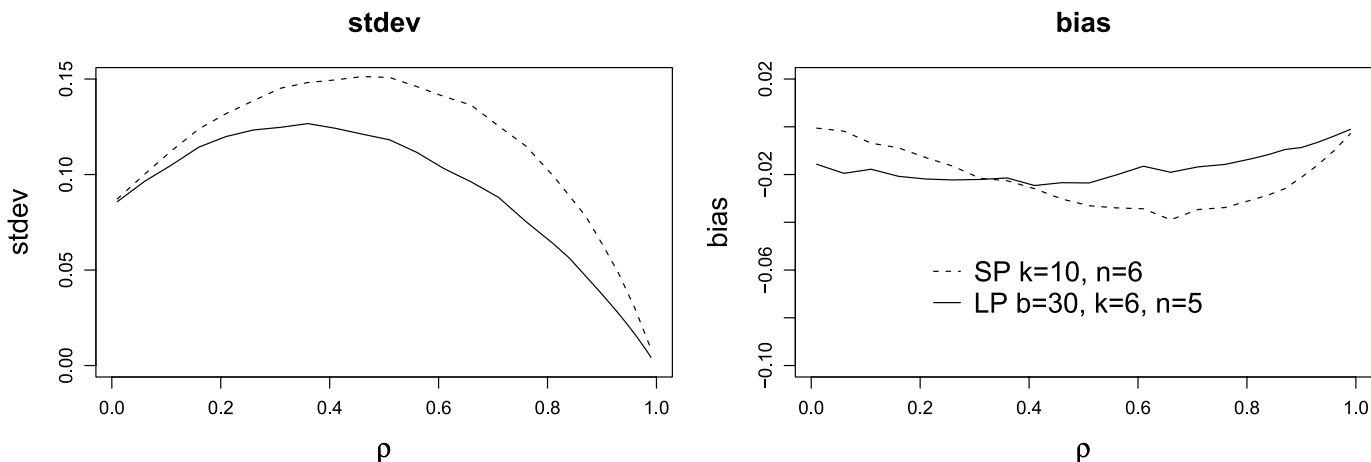

Figure 1. Comparison of the bias and standard deviation for estimators of ρ in a LP.

Figure 2. Comparison of standard deviation and bias for a leveraged and standard plan.

cific values of ρ, namely 0.80 and 0.91. The value $\rho = 0.91$ corresponds to the minimum acceptable value as specified in Automotive Industry Action Group (2002). We also include $\rho = 0.80$ to show how the standard deviation of $\hat{\rho}_c$ behaves with a poor measurement system.

When calculating the asymptotic variance (19) for the combined estimator, we need to replace $\frac{1}{SSC}$ by its expected value because σ_r^2, as shown in (15), depends on $E[\frac{1}{SSC}]$. This quantity corresponds to the sum of the standardized squares of the k chosen observations from the baseline. In an LP, we choose parts to be remeasured based on their extreme baseline values which correspond to the $k/2$ lowest and $k/2$ highest observed values in the baseline. For odd values of k, we chose the extra part with baseline value to be larger than the baseline average. This implies these extreme baseline values can be represented as order statistics from the standard normal distribution. We can write for k even,

$$E\left[\frac{1}{SSC}\right]$$

$$= E\left(\frac{1}{Z_{[1:b]}^2 + \cdots + Z_{[k/2:b]}^2 + Z_{[b-k/2+1:b]}^2 + \cdots + Z_{[b:b]}^2}\right),$$

$$(23)$$

where $Z_{[i:b]}$ is the ith order statistic from a sample of b standard normal random variables. We estimate (23) by simulating 10,000 samples of b observations.

To start, we consider $N = 60$. In Table 4, we give the approximate standard deviation, as given by (19), of the combined estimator for ρ for the five top, middle, and bottom LP plans (out of approximately 200 choices). The standard deviation for each design was calculated using (19). These results are very close to those obtained through simulation. Plans with higher standard deviations tend to have a low number of observations allocated to the baseline. The designs with the lowest standard deviation have $b \simeq nk$, that is, $b \simeq N/2$.

In Table 5, we show the plans corresponding to the lowest stdev($\hat{\rho}_c$) for different values of $N = b + nk$ when ρ equals 0.80 and 0.91. We see that as ρ increases, the optimal value of k decreases somewhat. The differences in performance among the plans with the same N is small, but notice that the baseline size b is close to $N/2$ for all the best plans. Using this empirical evidence, we suggest using $b \simeq N/2$, $n \simeq 5$ and then $k \simeq N/10$. This plan is in every set of the top five in Table 5. Since the LP design parameters must be integers, given a total sample size N, we recommend the plan with k equal to the greatest integer less than or equal to $N/10$, $n = 5$, and $b = N - 5k$.

In Table 6, assuming that we use the recommended plan, we give the value of N required to achieve a specified standard error of the transformed estimator (21) when given a value of ρ.
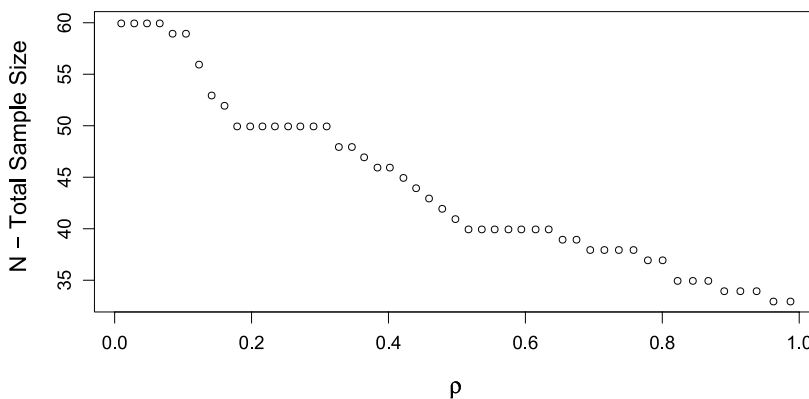

Figure 3. Sample size requirements for a leveraged plan with the same standard deviation as the standard plan.

Table 4. Estimation precision for ρ for a variety of LPs for $b + nk = 60$

	$\rho = 0.80$				$\rho = 0.91$			
	Baseline size b	# of parts k	# meas. per part n	stdev$(\hat{\rho}_c)$	Baseline size b	# of parts k	# meas. per part n	stdev$(\hat{\rho}_c)$
Top 5 plans	32	7	4	0.0684	32	4	7	0.0350
	30	6	5	0.0688	33	3	9	0.0351
	33	9	3	0.0688	30	5	6	0.0351
	30	10	3	0.0689	30	6	5	0.0352
	35	5	5	0.0690	30	3	10	0.0352
Middle 5 plans	18	14	3	0.0785	38	11	2	0.0394
	42	1	18	0.0785	45	3	5	0.0396
	22	2	19	0.0788	45	1	15	0.0397
	25	1	35	0.0789	21	1	39	0.0401
	18	7	6	0.0792	20	2	20	0.0401
Bottom 5 plans	7	1	53	0.1766	6	3	18	0.1017
	6	2	27	0.1831	6	2	27	0.1058
	5	5	11	0.1870	6	1	54	0.1138
	6	1	54	0.2053	5	5	11	0.1203
	5	1	55	0.2475	5	1	55	0.1496

Table 5. The five plans with the lowest stdev$(\hat{\rho}_c)$ for different values of $N = b + nk$

	$\rho = 0.80$					$\rho = 0.91$			
Total N	Baseline size b	# of parts k	# meas. per part n	stdev$(\hat{\rho}_c)$	Total N	Baseline size b	# of parts k	# meas. per part n	stdev$(\hat{\rho}_c)$
30	18	3	4	0.1065	30	18	2	6	0.0552
30	18	4	3	0.1068	30	16	2	7	0.0555
30	15	5	3	0.1076	30	18	3	4	0.0556
30	18	2	6	0.1078	30	18	1	12	0.0557
30	15	3	5	0.1081	30	17	1	13	0.0558
50	26	6	4	0.0766	50	26	4	6	0.0393
50	26	8	3	0.0769	50	26	3	8	0.0393
50	25	5	5	0.0770	50	29	3	7	0.0393
50	30	5	4	0.0771	50	28	2	11	0.0394
50	29	7	3	0.0771	50	25	5	5	0.0395
75	39	9	4	0.0599	75	40	5	7	0.0306
75	40	7	5	0.0601	75	39	6	6	0.0306
75	43	8	4	0.0602	75	39	4	9	0.0307
75	39	12	3	0.0603	75	40	7	5	0.0307
75	35	10	4	0.0603	75	43	4	8	0.0308
100	52	12	4	0.0507	100	51	7	7	0.0259
100	48	13	4	0.0509	100	52	6	8	0.0259
100	50	10	5	0.0509	100	52	8	6	0.0259
100	56	11	4	0.0509	100	50	10	5	0.0260
100	55	9	5	0.0510	100	50	5	10	0.0260
125	65	15	4	0.0447	125	62	9	7	0.0228
125	61	16	4	0.0448	125	65	10	6	0.0229
125	65	12	5	0.0448	125	61	8	8	0.0229
125	69	14	4	0.0449	125	62	7	9	0.0229
125	60	13	5	0.0449	125	65	6	10	0.0229
200	100	25	4	0.0346	200	102	14	7	0.0177
200	104	24	4	0.0347	200	98	17	6	0.0177
200	100	20	5	0.0347	200	95	15	7	0.0177
200	96	26	4	0.0347	200	96	13	8	0.0177
200	105	19	5	0.0347	200	104	12	8	0.0177

Table 6. Values of N for estimating ρ with a specified standard deviation for the recommended plan

stdev($\widehat{\theta}$)	Assumed value of ρ					
	0.2	0.4	0.6	0.8	0.91	0.99
0.25	22	27	32	39	44	49
0.20	31	38	45	55	62	69
0.15	48	60	73	89	101	113
0.14	54	68	82	101	115	127
0.13	62	77	94	115	131	146
0.12	71	89	109	133	152	168
0.11	83	105	128	157	178	198
0.10	98	125	153	188	213	236
0.09	120	154	188	231	261	289
0.08	151	194	238	292	329	362
0.07	197	256	314	383	429	469
0.06	273	356	436	528	586	633
0.05	409	538	657	780	852	908

For example, suppose historical data suggests $\rho \approx 0.91$ and we want to estimate ρ with a standard deviation of at most 0.025. Then from (22), we have stdev($\widehat{\theta}$) $\approx$ stdev($\hat{\rho}_c$) $\frac{1}{1-\rho^2}$. Thus, we require the standard deviation on the transformed scale to be $\frac{0.025}{1-0.91^2} \approx 0.145$. Now in Table 6, we look down the column with $\rho = 0.91$ and stdev($\widehat{\theta}$) = 0.15 to get the total sample size of 101. Using the suggested plan, we require 51 parts for the baseline study. Then, from the baseline, we select $k = 10$ extreme parts, five on each side of the average, to repeatedly measure them $n = 5$ times each.

6. MODEL ASSUMPTIONS AND CONSIDERATIONS

To use an LP, we recommend selecting a number of extreme parts in Stage 2. Questions then arise about the sensitivity of the LP (relative to a SP) to the model assumptions and to methods for detecting departures from the assumed model.

A key assumption of model (1) is that the properties of the measurement system are independent of those of the underlying process. This assumption is sometimes called linearity of the measurement system (Automotive Industry Action Group 2002). In particular, we are assuming that the bias (if any) and the standard deviation σ_m do not depend on the part size P. Many measurement systems are nonlinear in that σ_m increases as the part size increases.

If the bias is constant across part size, this bias gets subsumed into the process mean and has no effect on the estimation of ρ for either a LP or SP. If the bias varies across part size so that, given $P_i = p_i$, the mean of E_{ij} in (1) depends on p_i, then we can rewrite the model so that the part effect is redefined to include this bias. With either plan, we can estimate ρ but its value is inflated because the varying bias is included in the process variation. Neither the SP or LP can detect varying bias.

If the measurement variability is a function of part size, then, with either plan, the meaning of ρ is not clear and we are unsure what it is we are estimating. Since we repeatedly measure parts a number of times with either a SP or an LP, we have some power to detect if σ_m is varying over part size. With a LP, we typically have fewer repeated measurements. However, if the measurement variability is increasing with part size, by using extreme parts we have a greater chance to detect the difference. It is unclear which plan has the advantage to detect this type of nonlinearity.

For a LP, we can check the normality assumption for the baseline measurements using a QQ plot. We can also construct a QQ plot for the residuals of the repeated measurements, ignoring the baseline measurement, to check the normality of the measurement errors. In practice, if an outlier is observed in the baseline measurements, we do not recommend the use of the corresponding part in the second stage. Such an outlier may be due to either the process or the measurement system. We would advise a separate study of this part, because, if the extreme value is due to the measurement system, finding such an outlier in a small baseline study suggests that there may be a larger problem with the measurement system.

A referee raised the issue of robustness of the LP to departures from normality to underlying distributions with heavier tails. One might expect problems because of the use of parts with extreme baseline measurements. To investigate this issue, we conducted a small simulation with four cases: both the part and measurement error distribution are normal, the part distribution is t_5 and the measurement error distribution is normal, the part distribution is normal and the measurement error distribution is t_5, and both the part and measurement error distributions are t_5. From the simulation, we found that changing the part or measurement error distribution has surprisingly little effect other than a small increase in the standard deviation for all estimators and a change in the bias of the ANOVA estimate.

In summary, we can assess the assumptions of model (1) as easily as for the LP as for the SP. Departures from the model affect both plans, but there is no evidence, based on a cursory study, that the LP has greater sensitivity to these departures than does the SP.

7. DISCUSSION AND SUMMARY

For the purpose of this paper, we assume the total variation σ_t and the process mean μ are unknown but interest lies in estimating ρ. This assumption is reasonable for a new measurement system or for an existing measurement system where we do not use prior information and matches common practice. In Browne, Mackay, and Steiner (2009), we demonstrate the advantages of leveraging for the routine assessment of a measurement system currently used for 100% inspection where we can assume μ and σ_t are known so that σ_m is the only remaining unknown. For leveraging to be effective in this case, we must have a supply of premeasured parts.

There are widely used performance measures other than ρ such as the PT ratio, that is, $\sigma_m/tolerance$, that depend only on σ_m so there is no need to estimate μ and σ_t. If our assumptions are correct, the best plan for estimating σ_m is to repeatedly measure any single part. In this case leveraging does not help. However, if a supply of premeasured parts is available (effectively a baseline sample), then selecting the part with initial measurement far from the average increases the efficiency of estimation.

One extension to the proposed LP is to consider making a different number of measurements on the selected extreme parts.

We can show that by repeatedly measuring the more extreme parts more often that we can increase efficiency. The gain from this effort is marginal compared to the increased complexity of the plan.

Leveraging can be applied to other variance component problems in which one source of variation can be held fixed (e.g., the process) and the other can be varied (e.g., the measurement system). One example is an assembly–disassembly experiment (see Steiner and Mackay 2005, sec. 10.4) where we have an assembled product with several components. The goal is to determine which is the greater source of output variation, the components or the assembly operation. We select $k = 2$ products for the study. Each product is disassembled, reassembled, and then remeasured. These three steps are repeated several times. Performing the experiment in this way ensures the variation from the components is held fixed while the assembly contribution varies in each run. We use leveraging by selecting the two products for the experiment to have extreme and opposite performance when initially measured. To make this selection, we need a large number of previously measured products.

In summary, we present a new two-stage plan that uses leveraging to assess the intraclass correlation coefficient ρ of a measurement system. We define leverage to be the purposeful selection of parts with extreme baseline measured values. We show that a LP with the same number of total measurements is more efficient than the SP in which the parts to be remeasured are sampled randomly from the process. In terms of planning, we recommend a LP that uses about half the measurement effort to generate the baseline sample in Stage 1. This sample must be representative of the process to get reasonable estimates of the overall process mean and standard deviation. In Stage 2, we select k extreme parts from the baseline sample so that the average of their baseline measurements is close to the baseline average. We choose k sufficiently small so that we have resources to measure each of these parts $n = 5$ additional times.

ACKNOWLEDGMENTS

APPENDIX A: CONDITIONAL DISTRIBUTION

The conditional distribution given a baseline measurement does not depend on the rank of the baseline measurement from a sample.

Theorem 1. If $Y_{ij} = P_i + E_{ij}$ where $P_i \sim \mathrm{N}(0, \sigma_p^2)$ and $E_{ij} \sim \mathrm{N}(0, \sigma_m^2)$ $(i = 1, 2, \ldots, b$ and $j = 0, 1, 2, \ldots, n)$ then if we sample $\{Y_{10}, \ldots, Y_{b0}\}$ and order them such that $\{Y_{(1)0} \leq \cdots \leq Y_{(b)0}\}$ then the conditional distribution $Y_{(i)1}, \ldots, Y_{(i)n} | Y_{(i)0}$ is given in (8).

Proof. The distribution of $\{Y_{i0}, Y_{i1}, \ldots, Y_{in}\}$ is multivariate normal as given in (7).

From the properties of the multivariate normal distribution (Dillon and Goldstein 1984) we have that $Y_{i1}, \ldots, Y_{in} | Y_{i0} = y_{i0}$ is given by (8).

Rewriting the joint density of $\{Y_{10}, Y_{i1}, \ldots, Y_{1n}, Y_{20}, \ldots, Y_{2n}, \ldots, Y_{b0}, \ldots, Y_{bn}\}$ we get

$$h(y_{10}, \ldots, y_{bn})$$

$$= \prod_{i=1}^{b} f(y_{i0}, y_{i1}, \ldots, y_{in})$$

which the distribution defined in (7)

$$= \prod_{k=1}^{b} f(y_{k1}, \ldots, y_{kn} | y_{k0}) f(y_{k0})$$

which the distribution defined in (8)

$$= \prod_{k=1}^{b} f(y_{k1}, \ldots, y_{kn} | y_{k0}) \prod_{k=1}^{b} f(y_{k0})$$

do a change of variables such that

$$y_{(1)0} \leq y_{(2)0} \leq \cdots \leq y_{(b)0}$$

$$= \left[\prod_{k=1}^{b} f(.y_{k1}, \ldots, y_{kn} | y_{k0}) \right] n! \prod_{j=1}^{b} f(y_{(j)0})$$

integrate all y_{kl} where $k \neq i$

$$= f(y_{i1}, \ldots, y_{in} | y_{i0}) \frac{n!}{(i-1)!(n-i)!}$$

$$\times \left[F(y_{(i)0}) \right]^{i-1} \left[1 - F(y_{(i)0}) \right]^{n-i} f(y_{(i)0})$$

$$= f(y_{i1}, \ldots, y_{in} | y_{i0}) f(y_{(i)0}).$$

We can see that this is the joint distribution of $\{Y_{(i)0}, Y_{(i)1} \ldots, Y_{(i)n}\}$. Thus, the conditional distribution of $Y_{(i)1}, \ldots, Y_{(i)n} | Y_{(i)0}$ is (8).

APPENDIX B: FISHER INFORMATION

To show that $SC = 0$ and $SSC \gg 0$ reduces the asymptotic variance of the MLE, let

$$\mathbf{J}(\mu, \sigma_t^2, \rho) = \begin{pmatrix} x & 0 & t \\ 0 & y & v \\ t & v & z \end{pmatrix},$$

where $x, y, z \geq 0$. Using the principal minors, the determinant and inverse of $\mathbf{J}$ are

$$\det(\mathbf{J}) = x \begin{vmatrix} y & v \\ v & z \end{vmatrix} - 0 + t \begin{vmatrix} 0 & t \\ y & v \end{vmatrix} = x(yz - v^2) - t^2 y,$$

$$\mathbf{J}^{-1} = \frac{1}{\det(\mathbf{J})} \begin{pmatrix} yz - v^2 & vt & -yt \\ vt & xz - t^2 & -xv \\ -yt & -xv & xy \end{pmatrix}.$$

This means the asymptotic variance of maximum likelihood estimator of ρ is

$$\mathrm{Asvar}(\hat{\rho}) = \frac{xy}{x(yz - v^2) - t^2 y} = \frac{1}{z - v^2/y - t^2/x}.$$

Ideally, Asvar($\hat{\rho}$) is close to zero. From (10) we see that selecting parts to repeatedly remeasured affects t and z. Asvar($\hat{\rho}$) is reduced when z is large and $t = 0$. Since, x, y, $z \geq 0$, we can reduce Asvar($\hat{\rho}$) by decreasing v^2 or t^2. We cannot change v, but we can set $t = 0$ by selecting parts with baseline measurements such that $E[SC] = 0$. If we choose parts with large and small extreme measurements we can get $E[SC] = 0$, which means $t = 0$ and $E[SSC]$ is large which increases z.

[Received December 2007. Revised January 2009.]

REFERENCES

Automotive Industry Action Group (2002), *Measurement Systems Analysis* (3rd ed.), Southfield, MI: Chrysler, Ford, General Motors Supplier Quality Requirements Task Force.

Browne, R. P., Mackay, R. J., and Steiner, S. H. (2009), "Leveraged Measurement System Assessment for Processes With 100% Inspection," *Journal of Quality and Technology*, to appear.

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

Donner, A., and Eliasziw, M. (1987), "Sample Size Requirements for Reliability Studies," *Statistics in Medicine*, 6, 441–448.

Jorgensen, B., and Knudsen, S. J. (2004), "Parameter Orthogonality and Bias Adjustment for Estimating Functions," *Scandinavian Journal of Statistics*, 31, 93–114.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2001), *Introduction to Linear Regression Analysis*, New York: Wiley.

Steiner, S. H., and Mackay, R. J. (2005), *Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Processes*, Milwaukee, WI: ASQ Quality Press.

Steiner, S. H., MacKay, R. J., and Ramberg, J. S. (2008), "An Overview of the Shainin System for Quality Improvement" (with discussion), *Quality Engineering*, 20, 6–45.