Two Stage
Leveraged Measurement System Assessment

Ryan Browne, Jock MacKay and Stefan Steiner

Dept. of Statistics and Actuarial Sciences

Business and Industrial Statistics Research Group

University of Waterloo

Waterloo, N2L 3G1, Canada

rpbrowne@uwaterloo.ca

December 19, 2007

To assess measurement system variation, we propose an alternative to the traditional plan of using 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 study, a number of parts are measured once. In the second stage, we select a few parts based on their initial measurements from the baseline sample and remeasure them each a number of times. The new and traditional approaches are compared using the standard deviation of the estimators of the interclass correlation coefficient. We also present a method to determine sample size requirements for using the two stage Leveraged Measurement System Assessment.

Key Words: leverage, intraclass correlation, maximum likelihood

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 measurement systems.

In this article, we use manufacturing terminology but the proposed plan and analysis can be applied in a wide variety of settings. For example, we can replace part-to-part variability by person-to-person variability when assessing the measurement variation in a medical instrument.

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

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

where P_i is a random variable representing the possible values for the true dimension of part i, E_{ij} is a random variable representing the error on the j^{th} measurement, 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 $\sigma_t^2 = \sigma_p^2 + \sigma_m^2$. As in a traditional measurement assessment, by adopting model (1), we assume that μ, σ_p and σ_m are stable over the time needed to conduct the measurement assessment investigation and that σ_p and σ_m are constant across true part dimensions.

We also assume the studied measurement system is automated, that is there are no operator effects. Automated measurement systems are common. For instances, in one example, piston diameters were measured by an inline gauge with automated part handling. Using manufacturing jargon, σ_m captures measurement repeatability and not reproducibility.

In this article, we present a measurement system assessment study where we assume there is no prior information about the parameters in model (1). This plan is applicable when assessing a new measurement system or an existing measurement system where we do not use prior information. This matches common practice. In Browne et al. (2007) we looked at the case of the routine assessment of a measurement system currently used for 100% inspection of a high volume process where we could assume μ and σ_t were known.

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, i.e. $\rho = \sigma_p^2/\sigma_t^2$. The intraclass coefficient is equivalent to the gauge repeatability defined in our notation as σ_m/σ_p.

The common cut-off values (Automotive Industry Action Group, 2002) for gauge repeatability used to determine if a measurement system is acceptable are 0.10 and 0.30. These values correspond to ρ equal to 0.91 and 0.99 respectively.

The standard measurement assessment plan (Automotive Industry Action Group, 2002) is to sample k parts selected at random from the process and measure them n times each for a total of $N = nk$ measurements. This results in the data y_{ij}, where $i = 1, \ldots, k$ and $j = 1, \ldots, n$. Following Donner and Eliasziw (1987), for the standard plan, we estimate ρ by the sample intraclass correlation coefficient

$$\frac{\text{MSA} - \text{MSW}}{\text{MSA} - (n-1)\text{MSW}} \tag{2}$$

where MSA and MSW are

$$MSA = \sum_{i=1}^{k} n\left(\overline{y}_{i.} - \overline{y}_{..}\right)^2/(k-1) \qquad MSW = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(y_{ij} - \overline{y}_{i.}\right)^2/\left[k(n-1)\right]$$

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

In this article we compare this standard plan (SP) to a new Leveraged plan (LP) by comparing the standard deviation and bias of a number of estimators. A LP has two stages. In the first stage, called the baseline study, b parts are measured once. In the second stage, a sub-sample of k parts is chosen from the baseline sample using the observed measured values. These k parts are then repeatedly measured n times. The total number of measurements for the leveraged plan is $N = b + nk$.

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. The parameters μ and σ_t are viewed as nuisance parameters.

This paper is structured as follows. In the next section, we describe the analysis for a LP including properties of the Maximum Likilihood Estimator (MLE) for ρ and other simpler estimators. In Section 3, we compare the standard and leveraged plans using the bias and standard deviation of the derived estimators for ρ. In Section 4, we compare different designs (i.e. different values of b, k and n) for leveraged plans when the total sample size is fixed. Based on empirical evidence, we recommend specific plans for any total sample size. Finally, we provide some discussion and draw conclusions.

2 THE LEVERAGED MEASUREMENT ASSESSMENT PLAN

The leveraged measurement system assessment plan is conducted in two stages:

Stage 1: Sample b parts at random from the process to obtain a baseline. We denote the observed values $\{y_{10}, y_{20}, \ldots, y_{b0}\}$.

Stage 2: From the baseline sample, select k parts using the observed measured values. In particular, to improve our estimate for ρ, we sample k parts that are extreme relative to the baseline average. We denote the k selected parts using the set S. These k parts are then repeatedly measured n times each to give the additional data $\{y_{ij}, i \in S \ \ and \ \ j = 1, \ldots, n\}$. The total number of measurements in the leveraged plan is $N = b + nk$.

For example, for a leveraged plan with $k = 2$, we may pick the parts with the minimum and maximum initial measurement in the baseline sample.

We recommend repeatedly measuring the parts in Stage 2 over the range of conditions (time, environment, etc.) expected to capture the major sources of measurement variation. This recommendation matches the requirements for a standard measurement assessment study (Automotive Industry Action Group, 2002).

We present four approaches for estimation of ρ . The first uses Maximum Likelihood. The MLE has no closed form and must be found numerically. The other three methods estimate μ and σ_t^2 from the baseline information only and then estimate ρ using the repeated measurements, conditional on the baseline observations. The second estimate uses a regression approach since the

conditional mean of the repeated measurements depends on ρ. The third uses the variation within the repeated measurements to estimate ρ. Finally, the fourth estimator is a linear combination of the second and third estimators.

2.1 Likelihood

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

$$\begin{pmatrix} Y_0 \\ Y_1 \\ \vdots \\ Y_n \end{pmatrix} \backsim N\left(\mu \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \sigma_t^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & & \\ \vdots & & \ddots & \vdots \\ \rho & & \dots & 1 \end{bmatrix} \right). \tag{3}$$

The distribution of the repeated measurements $\{Y_1, \dots, Y_n\}$ on a single part given the initial measurement $Y_0 = y_0$ is

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \Bigg| Y_0 = y_0 \backsim N\left(\mu + \rho(y_0 - \mu) \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \Sigma = \sigma_t^2 \begin{bmatrix} 1-\rho^2 & & \rho(1-\rho) \\ & \ddots & \\ \rho(1-\rho) & & 1-\rho^2 \end{bmatrix} \right). \tag{4}$$

where $\sigma_t^2 = \sigma_p^2 + \sigma_m^2$ is the total variation. The covariance matrix Σ has a special form which allows us to obtain the following well known properties (Dillon and Goldstein, 1984):

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

Using the properties of Σ, we can write down 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 thus

$$\begin{aligned} l_r\left(\mu,\sigma_t^2,\rho\middle|\, y_{10},\ldots,y_{b0}\right) &= -\frac{nk}{2}\log\sigma_t^2 - \frac{nk}{2}\log(1-\rho) - \frac{k}{2}\log(1+n\rho) \\ &\quad -\frac{1}{2}\frac{1}{\sigma_t^2(1-\rho)(1+n\rho)}\left\{(1+n\rho)SSW + n\sum_{i\in S}\left[\overline{y}_{i.} - \mu - \rho(y_{i0}-\mu)\right]^2\right\} \end{aligned}$$

where y_{i0} is the baseline measurement for the i^{th} part, $\overline{y}_{i.}$ is the average of the repeated measurements for the i^{th} part and $SSW = \sum_{i\in S}\sum_{j=1}^{n}(y_{ij}-\overline{y}_{i.})^2$. Assuming the b parts in the baseline sample are selected at random from the process, the marginal log-likelihood of the baseline is

$$l_0(\mu,\sigma_t^2) = -\frac{b}{2}\log\sigma_t^2 - \frac{1}{2\sigma_t^2}\left\{SSB + b\left(\hat{\mu}-\mu\right)^2\right\}.$$

where $\hat{\mu} = \frac{1}{b}\sum_{i=1}^{b} y_{i0} = \overline{y}_{.0}$ is the average of the baseline measurements and $SSB = \sum_{i=1}^{b}\left(y_{i0}-\overline{y}_{.0}\right)^2$.

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

$$l(\mu,\sigma_t^2,\rho) = l_0(\mu,\sigma_t^2) + l_r\left(\mu,\sigma_t^2,\rho\middle|\, y_{10},\ldots,y_{b0}\right). \tag{5}$$

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 (5). Solutions can be found numerically.

We recommend choosing parts with extreme initial measurements because this decreases the asymptotic variance of the MLE of ρ. The asymptotic variance-covariance matrix of the maximum likelihood estimators is the inverse of the Fisher information matrix; given by

$$J\left(\mu,\sigma_t^2,\rho\right) = \begin{pmatrix} \frac{(1-\rho)nk}{\sigma_t^2(n\rho+1)} & 0 & \frac{nE[SC]}{\sigma_t(n\rho+1)} \\ 0 & \frac{1}{2}\frac{b+nk}{\sigma_t^4} & -\frac{1}{2}\frac{nk\rho(n+1)}{\sigma_t^2(n\rho+1)(1-\rho)} \\ \frac{nE[SC]}{\sigma_t(n\rho+1)} & -\frac{1}{2}\frac{nk\rho(n+1)}{\sigma_t^2(n\rho+1)(1-\rho)} & E\left[-\frac{\partial^2}{\partial\rho^2}\, l\left(\mu,\sigma_t^2,\rho\right)\right] \end{pmatrix}, \tag{6}$$

where

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

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

We show in Appendix B that if a sampling plan is chosen such that $E\left[SC\right] = 0$, then the asymptotic variance of the MLE for ρ is reduced. Also, we show that the variance of the MLE is reduced by choosing a sampling plan where $E\left[SSC\right]$ is large. A plan with both these properties is to choose an equal number of parts with extreme initial measurements on either side of the baseline average.

In practice, if a wild outlier is obtained in the baseline measurements, we suggest caution in using this part in the second stage of the analysis because it may not represent the typical distribution of the measurements. As with traditional measurement studies, we recommend repeated measurements on this wild part to determine if the outlying measurement is due to the measurement system or the process.

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.

2.2 Regression Estimator

An alternative to maximum likelihood that has a closed form can be derived using a regression model. From (4), the distribution of the average of the repeated measurements on a single part, given the initial measurement y_{i0}, is

$$\overline{Y}_{i.}\,|(Y_{i0} = y_{i0}) \backsim N\left(\mu + \rho(y_{i0}-\mu),\ \ \sigma_t^2(1-\rho)\left(\rho + \frac{1}{n}\right)\right) \tag{7}$$

The averages of the repeated measurement on different parts are mutually independent. Since in (7) 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 $\overline{Y}_{i.}$ also depends on μ but we use the baseline average $\overline{y}_{.0}$ to estimate this unknown.

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

$$\hat{\rho}_r = \frac{\sum_{i \in S} (\overline{y}_{i.} - \overline{y}_{.0})(y_{i0} - \overline{y}_{.0})}{\sum_{i \in S} (y_{i0} - \overline{y}_{.0})^2} \tag{8}$$

If we standardize each quantity in (8), the marginal distribution of $\tilde{\rho}_r$, the corresponding estimator of the regression estimate, depends only on ρ. The distribution of $\tilde{\rho}_r$, conditional on the baseline sample, is normal with mean

$$E[\tilde{\rho}_r | y_{10}, \ldots, y_{b0}] = \rho + \left[(\mu - \overline{y}_{.0}) \frac{\widehat{SC}}{\widehat{SSC}} \right] (1 + \rho) \tag{9}$$

and variance

$$Var[\tilde{\rho}_r | y_{10}, \ldots, y_{b0}] = \frac{\sigma_t^2 (1 - \rho)(1/n + \rho)}{\sum_{i \in S} (y_{i0} - \overline{y}_{.0})^2}$$

$$\text{where} \quad \widehat{SC} = \sum_{i \in S} \left[\frac{y_{i0} - \hat{\mu}}{\hat{\sigma}_t} \right] \quad \text{and} \quad \widehat{SSC} = \sum_{i \in S} \left[\frac{y_{i0} - \hat{\mu}}{\hat{\sigma}_t} \right]^2 \tag{10}$$

are the baseline estimates of SC and SSC as defined in (6) and $\hat{\mu} = \overline{y}_{.0}$ is the baseline average and $\hat{\sigma}_t$ is the baseline standard deviation ${\hat{\sigma}_t}^2 = \sum_{i=1}^{b} (y_{i0} - \overline{y}_{.0})^2 / (b - 1)$. The estimator $\tilde{\rho}_r$ has a small bias (conditionally) because we choose parts so that $\widehat{SC} \approx 0$ and $\widehat{SSC}$ is large, thus $\widehat{SC}/\widehat{SSC} \approx 0$ and $\overline{y}_{.0}$ will be close to μ since the baseline sample is selected at random from the process. Unconditionally, the estimate is unbiased up to $O\left(\frac{1}{b}\right)$ because we can replace $\overline{y}_{.0}$ with $\mu + O\left(\frac{1}{b}\right)$. The unconditional variance of $\tilde{\rho}_r$ is

$$\sigma_r^2 = Var(\tilde{\rho}_r) \approx (1 - \rho) \left(\rho + \frac{1}{n} \right) E \left[\frac{\sigma_t^2}{\sum_{i \in S} \left(Y_{i0} - \overline{Y}_{.0} \right)^2} \right] \tag{11}$$

because $\tilde{\rho}_r$ is approximately unbiased. We estimate $E\left[\frac{\sigma_t^2}{\sum_{i \in S} \left(Y_{i0} - \overline{Y}_{.0} \right)^2} \right]$ from the baseline observations with the inverse of $\widehat{SSC}$ as given by (10). Similar to the MLE, choosing parts to re-measure with extreme baseline measurements relative to the baseline average reduces the conditional variance of this estimator. Since the estimator is nearly unbiased, selecting extreme parts will be also reduce the unconditional variance.

Note that the regression based estimator uses the average of the repeated measurements to

estimate ρ but not the variability in the repeated measurements unlike the next estimator.

2.3 ANOVA Estimator

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

$$MSW = \frac{\sum_{i\in S}\sum_{j=1}^{n}\left(y_{ij}-\overline{y}_{i.}\right)^2}{k(n-1)} \tag{12}$$

is an estimate of σ_m^2. Since the baseline variation is an estimate of σ_t^2 and $\rho = \sigma_p^2/(\sigma_p^2+\sigma_m^2)$, by rearrangement, we obtain the estimate

$$\hat{\rho}_a = 1 - \frac{\text{MSW}}{\hat{\sigma}_t^2}. \tag{13}$$

Transforming the ANOVA estimator, we see that $(1-\tilde{\rho}_a)/(1-\rho)$ 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. We have

$$E(\tilde{\rho}_a) = 1-(1-\rho)\,\frac{b-1}{b-3} \quad = \rho\,\left(\frac{b-1}{b-3}\right) - \frac{2}{b-3} \tag{14}$$

$$\sigma_a^2 = Var(\tilde{\rho}_a) = (1-\rho)^2\frac{2\,(b-1)^2(k(n-1)+(b-1)-2)}{k(n-1)((b-1)-2)^2((b-1)-4)}. \tag{15}$$

Note that neither the regression nor the ANOVA estimates require that the parts selected to be re-measured in stage two of the LP be representative of the process.

2.4 Combined Estimator

An estimator which has a closed form and turns out to have similar properties to the MLE is a combination of the two estimators $\tilde{\rho_r}$ and $\tilde{\rho_a}$ as described in Sections 2.2 and 2.3 respectively. We can show that, given the baseline data, these two estimators are conditionally independent. Furthermore, as seen in (9), the estimator $\tilde{\rho_r}$ is nearly unbiased. Hence the marginal covariance of $\tilde{\rho_r}$ and $\tilde{\rho_a}$ is close to 0. In what follows, we ignore this covariance. The combined estimator is a

linear combination of $\tilde{\rho}_r$ and $\tilde{\rho}_a$ with weights selected to minimize the variance.

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

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

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

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

Parameters are estimated by setting the estimating function to zero and solving. Substituting the quantities in (15) for σ_a^2 and (11) for σ_r^2, we obtain the combined estimate ρ_c, as a root of the quadratic equation (18).

$$\left(v_F - E\left[\frac{1}{SSC}\right]\right)\rho_c^2 + \left(E\left[\frac{1}{SSC}\right]\left[\hat{\rho}_a - \frac{1}{n}\right] - v_F\left[1+\hat{\rho}_r\right]\right)\rho_c + \left(v_F\hat{\rho}_r + E\left[\frac{1}{SSC}\right]\frac{\hat{\rho}_a}{n}\right) = 0 \tag{18}$$

where $v_F = Var\left(F_{k(n-1),b-1}\right)$. As with the regression estimator, we estimate $E\left[\frac{1}{SSC}\right]$ from the baseline observations with the inverse of $\widehat{SSC}$ as given by (10).

In this case, the appropriate estimator is the smaller root because the larger root gives estimates of ρ which are greater than one and ρ is bounded between zero and one. Note that $\tilde{\rho}_c$ is not just a simple weighted average of the two previous estimators 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

$$\mathrm{Var}(\tilde{\rho}_c) \approx \frac{Var\left[\Psi_c(\rho)\right]}{E\left[\frac{\partial}{\partial\rho}\Psi_c(\rho)\right]^2} = \frac{\sigma_a^2\sigma_r^2}{(\sigma_a^2+\sigma_r^2)} \tag{19}$$

The asymptotic variance covariance matrix (see Jorgensen and Knudsen, 2004) of $\tilde{\mu},\ \tilde{\sigma}_t^2$, as

estimated from the baseline, and $\tilde{\rho}_c$, as given by solving (18), is

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

The variance of the combined estimator depends on ρ through σ_a^2 and σ_r^2. Through simulation, we noticed that the distribution of $\tilde{\rho}_c$ is skewed towards zero when ρ is close to one. Qualitatively, this occurs because $\tilde{\rho}_a$ (which has a skewed distribution) has more weight in this situation.

To create confidence intervals using estimators with skewed distribution, it is common to work on a transformed scale. A transformation that seems to perform well here is the Fisher z transform. We let

$$\theta = \frac{1}{2}\log\frac{1+\rho}{1-\rho} \qquad \text{and} \qquad \frac{\partial\theta}{\partial\rho} = \frac{1}{1-\rho^2} \tag{21}$$

then its variance is approximately

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

To create a confidence interval for ρ, we first create a confidence interval on the θ scale. Then use the transformation on the confidence interval limits to create a confidence interval for ρ. We provide an example of applying this procedure in the next section.

2.5 Numerical Example of Various Estimates for ρ based on Leveraged Plan

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 the large and small parts were selected.

In the example, a baseline of 100 parts were randomly selected from the process. The baseline

data, given as a difference from nominal, are shown in Table 1.

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 baseline mean $\overline{y}_{.0}$ is 0.540 and variance is $\hat{\sigma}_t^2$ is 25.865. The parts chosen to be repeatedly measured were parts 70 and 50 (i.e. S=$\{70, 50\}$), with baseline measurements -12.2 and 12.8 respectively. These two parts were measured 18 more times each. The repeated measurements are given in Table 2 and plotted in Figure 1. 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.

Table 2: Example of a Stage 2 Sample with 2 Extreme Parts Repeatedly Measured 18 Times Each

Part 70 $y_{70,0} = -12.2$							Part 50 $y_{50,0} = 12.8$					
-10.3	-11.1	-10.0	-12.2	-11.0	-11.1		10.9	13.2	12.8	12.6	12.7	14.1
-10.9	-10.0	-10.6	-11.4	-11.5	-11.1		12.9	13.1	12.0	13.3	12.6	13.4
-11.4	-10.7	-10.3	-11.4	-9.8	-11.5		12.0	12.9	11.7	11.8	12.2	14.1
$\overline{y}_{70,.} = -10.9$ $s_{70}^2 = 0.40997$							$\overline{y}_{50,.} = -12.7$ $s_{50}^2 = 0.68029$					

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

$$\widehat{SC} = -2.51 + 2.41 = -0.10 \quad \text{and} \quad \widehat{SSC} = 6.275 + 5.811 = 12.086$$

The maximum likehilhood estimate for $\left(\mu, \sigma_t^2, \rho\right)$ is $(0.551, 25.392, 0.97809)$.

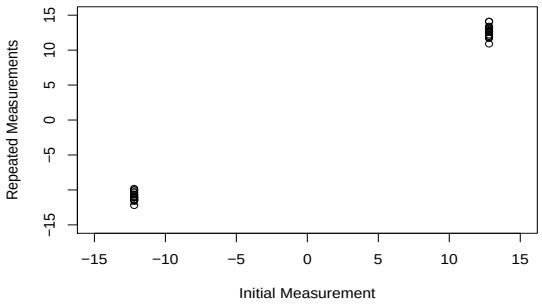

Figure 1: Plot of Repeated Measurements vs. Initial Measurement

Using (8), the regression estimate is

$$\hat{\rho}_r = \frac{\sum_{i \in S} \left(\overline{y}_{i.} - \overline{y}_{.0}\right)\left(y_{i0} - \overline{y}_{.0}\right)}{\sum_{i \in S} \left(y_{i0} - \overline{y}_{.0}\right)^2} = \frac{145.8 + 148.9}{162.3 + 150.3} = \frac{294.7}{312.6} = 0.94267$$

Using (12), the ANOVA estimate is

$$\hat{\rho}_a = 1 - \frac{MSW}{\hat{\sigma}_t^2} = 1 - \frac{(s_{50}^2 + s_{70}^2)/2}{\hat{\sigma}_t^2} = 1 - \frac{(0.40997 + 0.68029)/2}{25.865} = 0.97892$$

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 $v_F = Var\left(F_{34,99}\right) = 0.0845$ 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. Therefore 0.97816 is the combined estimate of ρ. Table 3 summarizes the four estimates and the corresponding standard errors. Note the very close agreement between the MLE and combined estimate.

To calculate confidence intervals for ρ, we work on the transformed scale since $\hat{\rho}$ is close to 1. We illustrate the calculations using the combined estimate. The transformed estimate, using (21) is $\frac{1}{2}\log\left(\frac{1+0.97816}{1-0.97816}\right) = 2.2531$. The standard error of this estimate is given in (22), using the standard

Table 3: Estimates of ρ for the Camshaft Journal Diameters Example

	Estimate	Standard Error
ρ_a	0.97892	0.00613
ρ_r	0.94267	0.06881
ρ_c	0.97816	0.00628
ρ_{mle}	0.97809	0.00597

error of the combined estimate (found in Table 3), works out to be $\frac{0.00628}{(1-0.97816^2)} = 0.14535$

A 95% confidence interval for θ is $2.25 \pm 1.96(1.45) = (1.968, 2.538)$ and the approximate 95% confidence interval in terms of ρ is $(0.962, 0.988)$.

2.6 Comparison of the Various Estimators for ρ in the Leveraged Plan

We consider a sampling plan with $b = 30$, $k = 6$ and $n = 5$ because this is the plan that we will recommend in Section 4 when the total number of measurements is 60. For stage 2 we suggest choosing the six parts which have the three largest and smallest measurements from the baseline study of 30 parts. Figure 2 shows the bias and standard deviation for the maximum likelihood estimator (MLE), regression estimator ($\tilde{\rho_r}$), ANOVA estimator ($\tilde{\rho_a}$) and the combined estimator ($\tilde{\rho_c}$). The figure was created by simulating ten thousand samples for each value of ρ in the range (0.01, 0.99).

The two individual components of the combined estimator, the regression and ANOVA estimators are efficient for different values of ρ. The standard deviation of the ANOVA estimator is much larger than the regression estimator when $\rho = 0.2$ but it performs well when ρ is larger than 0.9.

Notice that 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 estimator without loss of efficiency. We see similar results for leveraged sampling plans with other values of b, k and n.

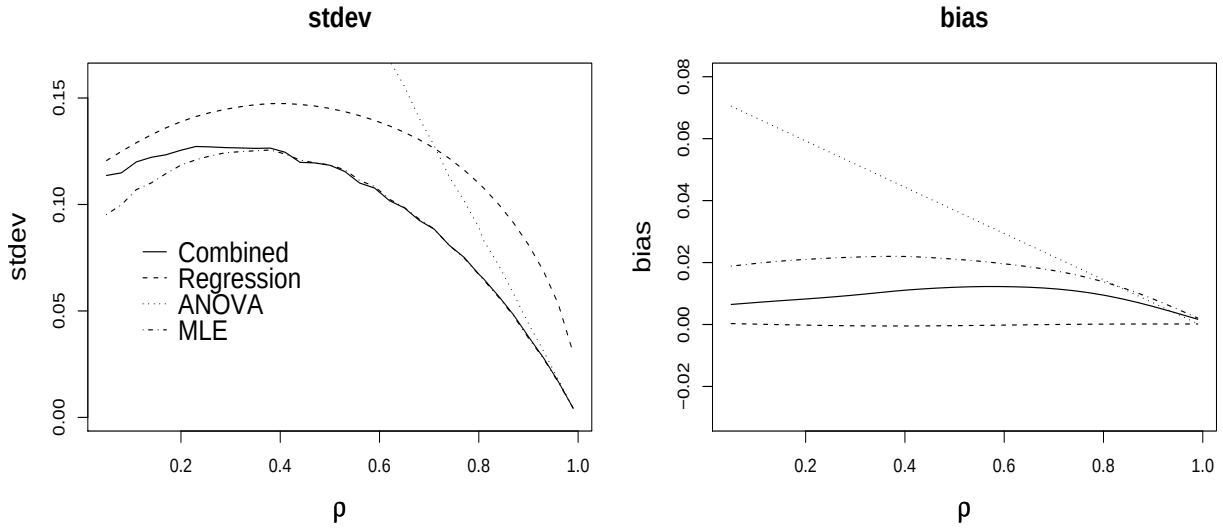

Figure 2: Comparison of the Bias and Standard Deviation for Estimators of ρ in a Leveraged Plan

3 THE VALUE OF LEVERAGING

To demonstrate the value of leveraging, we again resort to simulation. We will compare different version of the LP with the standard plan in which we chose $k = 10$ parts at random from the process and make $n = 6$ repeated measurements for each part, as recommended by the Automotive Industry Action Group (2002). We used maximum likelihood estimation in all cases to make the comparisons fair. We quantify the difference between the plans using bias and standard deviation calculated from 10,000 simulations at each value of ρ.

We compare the following two plans each with a total of 60 measurements

- SP with k =10 and n=6
- LP with $b = 30$, $k = 6$ and $n = 5$.

We see from Figure 3 that the LP is better than the SP. The LP has smaller standard deviation than the SP for all values of ρ and less bias when $\rho \geq 0.4$.

We conclude that leveraging improves the precision of the estimator without requiring more measurements. We can also compare the LP and SP by looking at the total number of measurements required to give a desired precision. Figure 4 shows the total sample size required for a LP to have

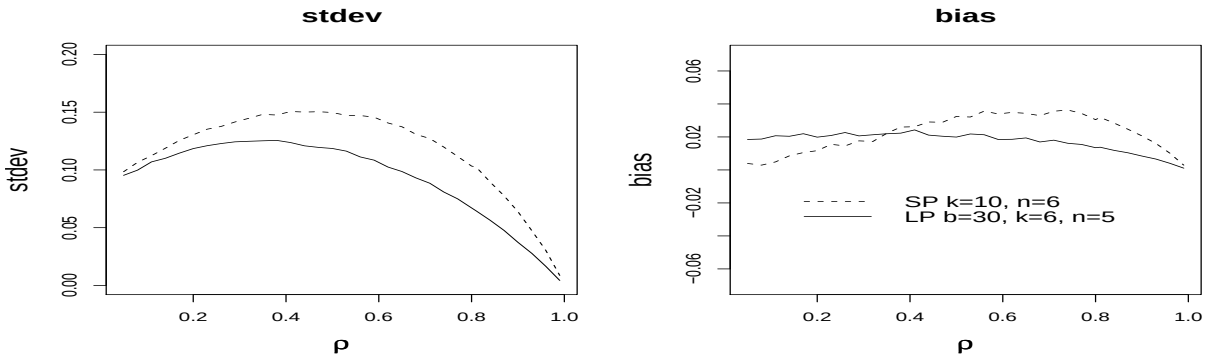

Figure 3: Comparison of Standard Deviation and Bias for a Leveraged and Standard Plan

the same precision (standard deviation) as the SP ($k = 10, n = 6$) for different values of ρ. In Figure 4, the selected LP corresponds to the suggested plan from Section 4 which has number of parts $k = \lfloor N/10 \rfloor$, number of repeated measurements per part $n = 5$ and a baseline of size $b = N - 5\lfloor N/10 \rfloor$, where $\lfloor \rfloor$ is the floor function that rounds down to nearest integer. For example, at $\rho = 0.91$, the SP has a standard deviation of 0.060 (see Figure 3). The LP with the same standard deviation for estimating ρ has a total sample size of 34, where $k = \lfloor 34/10 \rfloor = 3$, $n = 5$ and $b = 34 - 5\lfloor 34/10 \rfloor = 19$. We see similar results for other choices of n and k in the SP. We conclude that the two stage leveraged plan provides a substantial benefit.

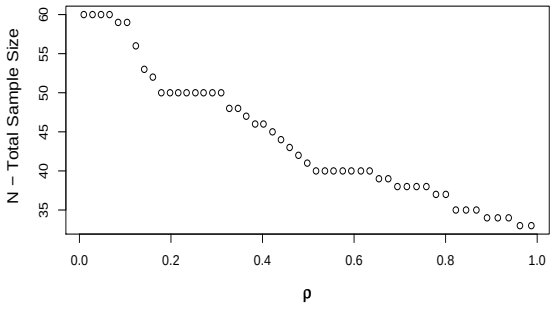

Figure 4: Sample Size Requirements for a Leveraged Plan with the same Standard Deviation as the Standard Plan

4 CHOICE OF LEVERAGED PLAN

In this section, we show how to choose a leveraged plan (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 is specified. As with most sample size calculations, we must also specify a value of ρ to select the plan. We consider two specific values of ρ, namely 0.80 and 0.91. These value of ρ are chosen because in terms of measurement repeatability they are equivalent to 0.45 and 0.30, respectively. Thus $\rho = 0.91$ corresponds to the minimum acceptable value in Automotive Industry Action Group (2002). We also include $\rho = 0.80$ in Tables 4 and 5 to show how the standard deviation of $\tilde{\rho}_c$ behaves with a poor measurement system.

When calculating the asymptotic variance (19) for the combined estimator, we need to replace $\frac{1}{\widehat{SSC}}$ by its expected value because σ_r^2, as shown in (11), depends on $E\left[\frac{1}{\widehat{SSC}}\right]$. This quantity corresponds to the sum of the standardized squares of the k chosen observations from the baseline. In a LP, we choose parts to be remeasured based on their extreme initial values which typically correspond to (assuming k is even) the $k/2$ lowest and $k/2$ highest observed values in the baseline. This implies these extreme initial values can be represented as order statistics from the standard normal distribution. We can write

$$E\left[\frac{1}{SSC}\right] = E\left(\frac{1}{Z^2_{[1:b]} + \cdots + Z^2_{[k/2:b]} + Z^2_{[b-k/2+1:b]} + \cdots + Z^2_{[b:b]}}\right) \tag{23}$$

where $Z_{[i:b]}$ is the i^{th} order statistic from a sample of b standard normal random variables. We estimate (23) by simulating ten thousand 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 $b \backsimeq nk$, i.e. $b \backsimeq N/2$.

In Table 5, we show the plans corresponding to the lowest stdev$(\tilde{\rho}_c)$ for different values of

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($\tilde{\rho}_c$)		Baseline Size b	# of parts k	# Meas. per part n	stdev($\tilde{\rho}_c$)
top 5	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	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	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

$N = b + nk$ when ρ equals 0.80 and 0.91. 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 $b \backsimeq N/2$, $n \backsimeq 5$ and then $k \backsimeq N/10$. This plan is in every set of the top 5 in Table 5. Since the LP design parameters must be integers, given a total sample size N, we recommend the plan with $k = \lfloor N/10 \rfloor$, $n = 5$ and $b = N - 5\lfloor N/10 \rfloor$.

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 variable (21) when given a value of ρ. We used the transformed scale because the distribution of the estimator is skewed when ρ is close to 1. 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($\tilde{\theta}$) $\approx$ stdev($\tilde{\rho}_c$) $\frac{1}{1-\rho^2}$. 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($\tilde{z}$) $= 0.15$ to get the total sample size of 101. Using the suggested plan, we require 51 parts for the baseline study. Then, from the baseline sample, we select 10 extreme parts to repeatedly measure 5 times each.

Table 5: The Five Plans with the Lowest stdev($\tilde{\rho}_c$) 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($\tilde{\rho}_c$)		Total N	Baseline Size b	# of parts k	# Meas. per part n	stdev($\tilde{\rho}_c$)
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

5 DISCUSSION AND CONCLUSIONS

In this paper, we present a new two stage plan for assessing the intraclass correlation coefficient of a measurement system using leveraging. We define leverage to be the purposeful selection of parts

Table 6: Values of N for Estimating ρ with a Particular Standard Deviation when $b = N - 5\lfloor N/10 \rfloor$, $n = 5$ and then $k = \lfloor N/10 \rfloor$

	Assumed value of ρ					
stdev($\tilde{z}$)	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

with extreme initial measured values to remeasured. We show that the leveraged measurement system assessment plan with the same number of total measurements is more efficient than the standard plan. Alternatively, to achieve the same precision in estimation for ρ as the SP, we can use a two stage LP with substantially fewer total measurements. We provide a closed form estimator for ρ that performs as well as the maximum likelihood estimator. We recommend a LP for a fixed number of total measurements N that has a baseline sample of size $b = N - 5\lfloor N/10 \rfloor$, we select a sample of the $k = \lfloor N/10 \rfloor$ most extreme parts from the baseline and we repeatedly measure each selected part $n = 5$ times.

As a possible extension to this work we can consider making a different number of measurements on the selected parts. Using Lagarange multipliers, we can show that by varying n_i on each part (in fact, increasing n_i for more extreme parts) we can increase efficiency but that the gain from this effort is marginal compared to the increased complexity of the plan.

Appendix A - Conditional Distribution

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

Theorem 1 *If* $Y_{i,j} = P_i + E_{i,j}$ *where* $P_i \sim N\left(0, \sigma_p^2\right)$ *and* $E_{i,j} \sim N\left(0, \sigma_m^2\right)$ $(i = 1, 2, \ldots, b$ *and* $j = 0, 1, 2, \ldots, n)$ *then if we sample* $\{Y_{1,0}, \ldots, Y_{b,0}\}$ *and order them such that* $\{Y_{1:b,0} \leq \ldots \leq Y_{b:b,0}\}$ *then the conditional distribution* $Y_{i:b,1}, \ldots, Y_{i:b,n} \mid Y_{i:b,0}$ *is given in (4)*

Proof. 1 The distribution of $\{Y_{i,0}, Y_{i,1}, \ldots, Y_{i,n}\}$ is multivariate normal as given in (3)

From the properties of the multivariate normal distribution (Dillon and Goldstein, 1984) we have that $Y_{i,1}, \ldots, Y_{i,n} | Y_{i,0} = y_{i,0}$ is given by (4)

Rewriting the joint density of $\{Y_{1,0}, Y_{i,1}, \ldots, Y_{1,n}, Y_{2,0}, \ldots, Y_{2,n}, \ldots, Y_{b,0}, \ldots, Y_{b,n}\}$ we get

$$
\begin{aligned}
& h\left(y_{1,0}, \ldots, y_{b,n}\right) \\
= & \prod_{i=1}^{b} f\left(y_{i,0}, y_{i,1}, \ldots, y_{i,n}\right) \quad \text{which the distribution defined in (3)} \\
= & \prod_{k=1}^{b} f\left(y_{k,1}, \ldots, y_{k,n} | \, y_{k,0}\right) f\left(y_{k,0}\right) \quad \text{which the distribution defined in (4)} \\
= & \prod_{k=1}^{b} f\left(y_{k,1}, \ldots, y_{k,n} | \, y_{k,0}\right) \prod_{k=1}^{b} f\left(y_{k,0}\right) \\
& \text{do a change of variables such that } \; y_{1:b,0} \leq y_{2:b,0} \leq \ldots \leq y_{b:b,0} \\
= & \left[\prod_{k=1}^{b} f\left(y_{k,1}, \ldots, y_{k,n} | \, y_{k,0}\right)\right] n! \prod_{j=1}^{b} f\left(y_{j:b,0}\right) \\
& \text{integrate all } y_{k,l} \text{ where } k \neq i \\
= & f\left(y_{i,1}, \ldots, y_{i,n} | \, y_{i,0}\right) \frac{n!}{(i-1)!(n-i)!} \left[F\left(y_{i:b,0}\right)\right]^{i-1} \left[1 - F\left(y_{i:b,0}\right)\right]^{n-i} f\left(y_{i:b,0}\right) \\
= & f\left(y_{i,1}, \ldots, y_{i,n} | \, y_{i,0}\right) f\left(y_{i:b,0}\right)
\end{aligned}
$$

We can see that this is the joint distribution of $\{Y_{i:b,0}, Y_{i:b,1} \ldots, Y_{i:b,n}\}$. Thus, the conditional distribution of $Y_{i:b,1}, \ldots, Y_{i:b,n} \mid Y_{i:b,0}$ is (4).

Appendix B - Fisher Information

To show that SC=0 and SSC≫0 reduce the asymptotic variance of the MLE, let

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

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

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

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

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

$$Asvar\left(\tilde{\rho}\right)=\frac{xy}{x(yz-v^2)-t^2y}=\frac{1}{z-v^2/y-t^2/x}$$

Ideally, $Asvar\left(\tilde{\rho}\right)$ is close to zero. From (6) we see that selecting parts to repeatedly remeasured affects t and z. The $Asvar\left(\tilde{\rho}\right)$ is reduced when z is large and $t=0$. Since, $x,y,z\geq 0$, we can reduce $Asvar\left(\tilde{\rho}\right)$ by decreasing v^2 or t^2. We cannot change v, but we can set $t=0$ by selecting parts with initial measurements such that $E\left[SC\right]=0$. If we choose parts with large and small extreme measurements we can get $E\left[SC\right]=0$, which means $t=0$ and $E\left[SSC\right]$ is large which increases z.

References

Automotive Industry Action Group (2002), *Measurement Systems Analysis* (Third ed.): Southfield MI.

Browne, R. P., Mackay, R. J., and Steiner, S. H. (2007), "Leveraged Measurement System Assessment," *Submitted to Technometrics*.

Dillon, W. and Goldstein, M. (1984), *Multivariate analysis methods and applications*: John Wiley & Sons.

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 Vinning, G. G. (2001), *Introduction to Linear Regression Analysis*: John Wiley & Sons.

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