Transpn. Res.-B, Vol. 26B, No. 2, pp. 135-154, 1992
Printed in Great Britain.

0191-2615/92 $5.00+.00

HIGHWAY PAVEMENT DISTRESS EVALUATION: MODELING MEASUREMENT ERROR

FRANNIE HUMPLICK
Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge MA 02139, U.S.A.

(*Received* 20 *April* 1990; *in revised form* 6 *February* 1991)

Abstract—There has been a proliferation of inspection technologies to quantify distresses on highway pavement systems. These technologies employ varying measurement principles and are subject to measurement errors. Estimates of measurement errors are therefore required in order to select among these techniques, and to get accurate assessments of pavement condition. There is abundant literature concerning techniques available for the numerical study of measurement errors. Available techniques include econometric methods of coping with errors in variables when investigating relationships between variables that have been measured with error; calibration approaches under various measurement conditions; and evaluation of measurement errors due to quantification, by a proxy, of concepts that are not directly measurable or observable. The methodologies employed in these techniques rely on certain assumptions for model development and estimation. These assumptions include the nature of error occurrence, whether systematic or random; the effect of these errors on the measured result, whether multiplicative or additive; and the level of knowledge about the true value of the measured object. Such assumptions may be violated under certain conditions. This paper identifies such situations and develops a generalized measurement error modeling approach, in which existing approaches are special cases. Existing methods are reviewed as to the specification models used to represent measurement errors; the types of errors accounted for in the suggested specification model; the type of errors not accounted for and the biases induced in estimated parameters by ignoring certain error types. The approach developed is capable of quantifying the accuracy of measurement for cases where the true value of the measured object is not known. This is the case in highway pavement distress evaluation as there is no single well-accepted technology which can be used as a proxy for the true value for calibration purposes. The methodology developed in this paper explicitly estimates the true value and is applied to the calibration of new technologies for highway distress evaluation.

1. INTRODUCTION

The study of measurement errors is a field of growing importance due to recent technological advances in data gathering and analysis. These technologies are now commonplace in evaluating distresses on highway pavements and include photographic, video and laser technologies. There are a variety of approaches to measurement error modeling such as econometric methods of coping with errors in variables; calibration techniques for measurement technologies that estimate biases and imprecisions of measurement under various conditions; and latent variable modeling techniques for dealing with measurement errors due to quantification of concepts that are not directly measurable or observable, but can be represented by a proxy.

Additionally, there is abundant literature concerning techniques available for the numerical study of errors. The methodologies employed in these techniques rely on certain assumptions for model development and estimation which may be violated if certain types of errors are present. This paper identifies such situations. The review of methods available in the literature is centered on four aspects:

(1) the model specification to represent measurement errors;
(2) the types of errors accounted for in the model specification;
(3) the types of errors not accounted for and the biases induced in estimated parameters by ignoring certain error types;
(4) suitability as a specification for the study of measurement errors in the context of distress evaluation on highway pavements.

The literature reviewed can be classified into two basic categories of measurement error analysis: (1) methods that examine existing processes and identify the presence or absence of errors, and (2) methods that determine the cause of error.

TR(B) 26:2-D

The first category of methods aim at detecting the presence of errors by studying the difference between two measurements of the same thing by two alternative means. The form that the difference takes, whether additive or multiplicative, is examined. These methods examine groups of measurements and detect differences between them, estimate the sizes and signs of these differences, and test their significance. Statistical comparison and calibration of measuring instruments, as well as correlation analysis, fall in this category.

The second category of methods aim at finding causality between observed differences in measurements by determining the factors responsible for the differences between measurements detected using the first category of methods. These methods identify the attributes of these factors that can be controlled. Such methods compare differences between measurements and find explanations for them, by identifying variables that are responsible for the difference. Experimental and quasi-experimental techniques, and analysis of variance and covariance, fall in this category.

The success of the second category of methodologies is dependent on the capability of the first category of methods to detect differences between measurements. The methodology developed in this paper is a hybrid of these two categories. It is a factor analytic specification which uses the covariance structures of multiple measurements to estimate biases. The techniques available in the literature are all special cases of the factor analytic representation. The following section describes the various methods available in the literature.

2. ERROR ANALYSIS METHODS AND TECHNIQUES

A review of some of the different situations in which measurements are made in the context of highway surface distress evaluation is introduced prior to discussing the state-of-the-art error analysis techniques. The quantity measured is the level of distress on a surface, which may be sparsely or densely distributed, systematically or haphazardly occurring on the surface, and visible to a greater or lesser extent. Sometimes, a proxy for the level of distress is measured, either because we cannot measure the level of distress directly or we do not know how to measure it.

Turning to the technologies used for measurement, they may automatically measure and record or semi-automatically record, or the entire operation may be conducted manually. The technologies, therefore, have varying levels of complexity and subjectivity. Errors of measurement can arise from the measurement system or inspector (human or machine), the inspected objects (distresses on various surfaces), or from the processing of data collected during information transmission and analysis, and during interpretation of the data collected.

The discussion above suggests that a variety of models may be needed to realistically describe the types of error relevant to the highway pavement condition inspection problem. Since the more error terms one wants to include, the more complex the formulation becomes, the objective is to obtain the simplest possible model that will reasonably represent the types of error of major interest. There are a number of simple models that exist in the literature. The assumptions inherent in these models, however, have not been thoroughly checked, and the consequences of these assumptions have not been examined in full. While there is abundant literature on methods of detecting large errors, or estimating the variances in measurements by different methods, little work has been done on estimating the types of error discussed above.

Specification alternatives

Two categories of methods for analyzing measurement errors can be distinguished: (i) statistical comparison and calibration; and (ii) experimental or quasi-experimental techniques. These methods require certain specification assumptions to represent measurement biases for different situations. The types of assumptions generally made are summarized below.

Consider the measurement of $k = 1, \ldots, K$ items on $i = 1, \ldots I$ locations by $j = 1, \ldots J$ technologies. This can be represented as shown below:

$$d_{ijk} = f(d^*_{ik}, \theta_{ijk}, \epsilon_{ijk}), \tag{2.1}$$

where:

d_{ijk} = measured level of item k in location i by a technology $j = 1, \ldots J$,
d^*_{ik} = true level of item k in location i, which is unknown,
θ_{ijk} = item, technology, and location- dependent measurement biases, that are also unknown,
ϵ_{ijk} = random error of measurement.

In this paper, the word "item" will be used to describe distress occurrences, and "location" to describe the sections on which the distress items occur. None of the terms on the right-hand side of (2.1) are observable. A number of assumptions about the bias parameters θ_{ijk}, and the true value d^*_{ik} are required for estimation. For ease of comparison and interpretation, linear representations of the function $f(.)$ will be used in the following discussions. The assumptions most commonly used and their interpretations are summarized below:

(1) Biases are both technology and item dependent. Each technology has its own capability to measure items and each item has its own distinct properties that affect how detectable and measurable it is to a technology. This assumption is valid for situations where each technology used has its own measurement principle, limitations and measurement strategy. These technological factors are affected differently by the properties of the items measured, so that a particular item characteristic may affect the measurement response of one technology and not of another. For example, the pattern of distress occurrence on a pavement surface may affect a technology that employs systematic sampling strategies and may not affect one that inspects the entire pavement surface. This specification can be represented by:

$$d_{ijk} = \alpha_{jk} + \beta_{jk} d^*_{ik} + \epsilon_{ijk} \tag{2.2}$$

where: α_{jk}, β_{jk} = additive and multiplicative components, respectively, of the bias parameters θ_{ijk}.

Studies recognizing that measurement errors were both item and technology dependent, date back to 1902, where an experiment known as the Pearson Experiment was conducted (Cochran, 1968). This experiment was an investigation of the nature of the errors of measurement when the quantity measured is fixed, and definite, and the measurements are taken by a human being. This case is less challenging than the situation this paper is addressing which involves the investigation of errors in infrastructure inspection, where the quantities measured are highly variable and ill-defined. However, the observations made by Pearson are very relevant for the infrastructure case, and the main findings are summarized below:

(a) The additive measurement bias varies from technology to technology and item to item.
(b) Random errors of measurement are positively correlated, that is the errors of two apparently independent observers in measuring the same quantity are positively correlated. This can be stated as:

$$E(\epsilon_{ijk}, \epsilon_{ij'k}) \geq 0, \qquad j \neq j'.$$

This is explained by the fact that factors affecting the result of measurement by one observer persist when the second observer is measuring. In the context of infrastructure surface inspection, this situation occurs when technologies applying the same measurement principle are used to measure the same distressed sections (e.g. two photographic techniques measuring cracking on the same group of pavement sections will be expected to have errors that are positively correlated). To remove the effect of this correlation, a randomization procedure assigning the sections to the technologies is required. Alternatively, one can select a group of technologies with radically different measurement processes to inspect the same sections, in the hope that the factors affecting technologies with different measurement processes are not the same.

To incorporate the Pearson results in the specification of a measurement equation requires techniques that introduce the effect of covariance between measurements by different technologies. Such techniques have been well studied in interview surveys where there are response biases due to the interviewee or from the interviewer (Fellegi, 1964; Hanson and Marks, 1958; and Hansen, Hurwitz and Bershad, 1961). Using the notation introduced earlier, this can be represented as follows:

$$d_{ijk} = \alpha_j + a_{jk} + \beta_{jk} d^*_{ik} + \epsilon_{ijk}. \tag{2.3}$$

The model in (2.3) is a variation of model (2.2) where the additive bias has two components: an overall technological bias α_j and a variable bias that depends on the technology j and the item measured k (denoted by a_{jk}). Special cases where the overall technological bias is assumed negligible ($a_j = 0$) can also be specified.

If a group of technologies measuring different distress types on different sections are used, the total variation in the results of measurement can be partitioned into the variation due to the technologies, distress types and sections using analysis of variance techniques. Such techniques have been used to estimate the variance components specifically due to the technologies and items measured. These include randomization alternatives that assign items to technologies to reduce or eliminate the variable bias components (see for example, Kish, 1965). These are discussed in detail in the following section.

(2) Biases are exclusively technology dependent. Technologies differ in their ability to detect and measure items, and this ability is independent of the items measured. This is a special case of the linear model described in eqn (2.2), and can be used for situations where one does not know in advance what type of item will be found on a location, that is the items (distresses) occur randomly. Thus, even if technologies are affected differently by different items, these effects will occur randomly, and will be captured by the random term ϵ_{ijk},

$$d_{ijk} = \alpha_j + \beta_j d^*_i + \epsilon_{ijk}, \tag{2.4}$$

where: α_j, β_j = technology specific measurement biases. This specification requires that the measurements d_{ijk} be in the same units. Equation (2.4) is also valid when the technologies have been randomly assigned to items as in the case of experimentally designed measurement plans.

(3) Biases are exclusively item dependent. Technologies are equally efficient in detecting and measuring items, but the items have distinct properties that affect how accurately they can be measured. This property can be referred to as detectability and measurability, and is a function of the physics of the measurement environment, which can be expressed by the characteristics of the items. These characteristics, which can be different for each item, affect the measurement response of a technology. For example, characteristics such as the contrast between a distress and the background on which it occurs affect how detectable a distress is to a photographic technique. This can be represented by:

$$d_{ijk} = \alpha_k + \beta_k d^*_{ik} + \epsilon_{ijk}, \tag{2.5}$$

where: α_k, β_k = item-dependent measurement biases.

The model in (2.5) assumes that all technologies have the same bias parameters, α_k, β_k and differ only in their random errors ϵ_{ijk}. The use of this model is justified on the basis that all measurement technologies evaluated are designed to detect and measure the same items, and how well they perform depends on how different the properties of the items they encounter in application are from the properties they were designed to measure. Any differences specific to a technology are assumed to be captured by the random term ϵ_{ijk}.

(4) All measurements are unbiased but imprecise. This alternative specification assumes that a number of technologies are used to measure the same items, and that the measurement biases are additive and item specific, but there is no overall measurement bias (Jaech, 1976). This is an *a priori* restriction on the general model in eqn (2.2). This restriction is justified when the set of technologies used, say $j = 1, \ldots, T$, have additive errors that tend to cancel out, that is, if one can select a group of technologies T, whose average measurements result in no overall bias of measurement. This is mathematically derived below.

Recall the relationship in eqn (2.2), assuming there is no multiplicative bias, i.e. $\beta_{jk} = 1$. It can be written as:

$$d_{ijk} = \alpha_{jk} + d^*_{ik} + \epsilon_{ijk}. \tag{2.6}$$

The assumption that for a set of technologies the net bias is zero means that:

$$\frac{1}{T}\sum_{j=1}^{T} \alpha_{jk} = 0.$$

Thus:

$$\overline{d_{ik}} = \frac{1}{T}\sum_{j=1}^{T} d_{ijk} = d^*_{ik} + \overline{\epsilon_{ik}}, \tag{2.7}$$

where:

$$\overline{\epsilon_{ik}} = \frac{1}{T}\sum_{j=1}^{T} \epsilon_{ijk}.$$

The relation in (2.7) is used to estimate the precision of the technologies $j = 1, \ldots, T$. This is performed by assuming that the locations i are selected at random and the $\bar{\epsilon}_{ik}$ are normally distributed with zero mean and variance σ_k^2. The assumption of zero net bias can be justified in the case of measuring distresses on infrastructure surfaces when multiple technologies that have radically different measurement principles are used. For example, human, photographic and laser inspection.

This model has been used very frequently in the literature (Grubbs, 1948; Grubbs, 1973; and Cochran, 1968). Estimators for the precision and large sample tests of hypotheses about these estimators have been developed for cases where two or more instruments are used (Jaech, 1976; Shukla, 1972; and Maloney and Rastogi, 1970).

The types of error that can be estimated by such a model are random errors of measurement. Systematic biases due to influences such as the inherent variability within the items or the technologies are not accounted for. Therefore, such a specification has limited use. The next section discusses the estimation alternatives that can be used.

Estimation alternatives: Regression approaches

A number of models were discussed in the previous section. These models were specified by making certain assumptions about the bias parameters. Various estimators for the additive bias α_j and the variance of the random term Var (ϵ_{ijk}) have been developed in the literature by assuming that the technologies have no multiplicative measurement bias, such as (Grubbs, 1948). Other statistical comparison and calibration techniques have been used for estimating the biases and precision of two or more technologies measuring the same items, or one technology measuring multiple items (Cochran, 1968; Grubbs, 1973; and Jaech, 1976).

Two assumptions can be made, which lead to the selection of different estimators of the technological parameters α_j,β_j, and Var(ϵ_{ijk}). The first assumption is that there is a technology that is unbiased whose measurements can be used as a replacement for the true value of an item, or one can develop an unbiased reference from an average of multiple technologies. The second assumption is a pre-restriction of the presence of certain error types, such as assuming the multiplicative bias is not present. The most general specification in the previous section (biases are both technology and item dependent) is used to describe the estimation approaches available.

Let us consider the generalized model in eqn (2.2) where the biases of measurement are due to both the technologies used and the distresses measured. For simplicity, let us consider the case where a single item is measured by multiple technologies:

$$d_{ij} = \alpha_j + \beta_j d_i^* + \epsilon_{ij}. \tag{2.8}$$

The relationship for a reference ("correct") technology can be written as follows:

$$d_{ir} = \alpha_r + \beta_r d_i^* + \epsilon_{ir}. \tag{2.9}$$

Rearranging (2.9) gives:

$$d_i^* = \frac{d_{ir} - \alpha_r - \epsilon_{ir}}{\beta_r}.$$

Substituting in eqn (2.8) gives:

$$d_{ij} = \alpha_j + \beta_j \frac{(d_{ir} - \alpha_r - \epsilon_{ir})}{\beta_r} + \epsilon_{ij},$$

$$d_{ij} = \left(\alpha_j - \frac{\beta_j}{\beta_r}\alpha_r\right) + \frac{\beta_j}{\beta_r}d_{ir} + \left(\epsilon_{ij} - \frac{\beta_j}{\beta_r}\epsilon_{ir}\right).$$

Let:

$$\begin{aligned} \alpha_j' &= \alpha_j - \frac{\beta_j}{\beta_r}\alpha_r, \\ \beta_j' &= \frac{\beta_j}{\beta_r}, \\ \epsilon_{ij}' &= \epsilon_{ij} - \frac{\beta_j}{\beta_r}\epsilon_{ir}. \end{aligned} \tag{2.10}$$

Then,

$$d_{ij} = \alpha_j' + \beta_j' d_{ir} + \epsilon_{ij}'. \tag{2.11}$$

The relationship in (2.11) is what is used in the literature to calibrate measurements by fallible technologies to those by a reference. For example, in a study conducted by the Iowa Department of Transportation (Iowa DOT) to evaluate automated pavement data collection equipment, visual inspection by Iowa DOT inspectors was used as a reference against which the photographic inspection system developed by PASCO was evaluated (Jeyapalan, Cable and Welper, 1987). Calibration models of the form in (2.11) were estimated for three distress types: rutting, cracking and patching. The estimated models had very poor fit with R^2 ranging from 0.11 for cracking, 0.21 for rutting and 0.44 for patching. The limitations of such a specification which lead to the poor fit can be demonstrated by the following explanation.

If the reference technology is truly unbiased, then $\alpha_r = 0$, and $\beta_r = 1$. The relationships in (2.10) become:

$$\begin{aligned} \alpha_j' &= \alpha_j, \\ \beta_j' &= \beta_j, \\ \epsilon_{ij}' &= \epsilon_{ij} - \beta_j\epsilon_{ir}. \end{aligned} \tag{2.12}$$

Then, eqn (2.11) becomes:

$$d_{ij} = \alpha_j + \beta_j d_{ir} + (\epsilon_{ij} - \beta_j\epsilon_{ir}). \tag{2.13}$$

The independent variable d_{ir} in eqn (2.13) is correlated with the error term $(\epsilon_{ij} - \beta_j\epsilon_{ir})$ because d_{ir} is correlated with ϵ_{ir} as shown in eqn (2.9). This is also true for eqn (2.11), because d_{ir} is correlated with ϵ_{ij}'. Therefore, any ordinary least squares estimation technique to obtain estimates of α_j and β_j or α_j' and β_j' will result in biased estimates. Specifically, the multiplicative bias parameters β_j or β_j' obtained through ordinary least squares (OLS) will be *underestimated*. This is because d_{ir} is positively correlated with ϵ_{ir}, and for most cases β_j is greater than zero, which implies that d_{ir} and ϵ_{ij}' are negatively correlated. This gives a $\hat{\beta}_{OLS}$ which is less than the true β, irrespective of the sample size used (see for example, Pindyck and Rubinfeld, 1981). Likewise, the estimated $\hat{\alpha}_{OLS}$ will be a biased and inconsistent estimator of the true additive bias α, while the estimated standard error of the regression and the standard errors will also be biased and inconsistent. Thus any decisions based on these estimates, may be erroneous. For example, since β_j will be underestimated, and since for most technologies β is less than one, this will lead to the interpretation that, technology j, has a larger multiplicative bias than it really has, as the more the β deviates from one, the larger the multiplicative bias.

Model specifications that require the assumption that there is an unbiased technology, result in biased estimates of the measurement errors if the errors are technology dependent. Any tests on the significance of parameters or tests on evidence of factors causing error based on such estimates are also biased.

A discussion of a number of alternative ways of getting around this problem suggested in the literature follows.

Estimation alternatives: Differencing techniques

The second assumption commonly made relates to the absence of multiplicative errors. Models that allow the estimation of systematic biases without requiring a reference technology have been developed by Hahn and Nelson (1970). For identification purposes, such models require at least three independent measurements. The type of measurements used in this approach consist of one reading from one technology and two additional independent readings from a second technology, on each of a number of items. They can be generalized to multiple technologies.

Estimates of the difference in means are obtained from subtracting the average measurements by the two technologies. An estimate of the error variance (precision) for the second technology is obtained from the average measurements with that technology. The Hahn and Nelson approach is demonstrated below:

Assume technology $j = 1$ measures a single distress item $k = 1$, and technology $j = 2$ measures two distress items $k = 1$ and 2. Then:

$$\begin{aligned} d_{i11} &= \alpha_{11} + d_{i1}^* + \epsilon_{i11}, \\ d_{i21} &= \alpha_{21} + d_{i1}^* + \epsilon_{i21}, \\ d_{i22} &= \alpha_{22} + d_{i2}^* + \epsilon_{i22}. \end{aligned} \tag{2.14}$$

Taking the differences between the pairs of equations in (2.14) gives:

$$
\begin{aligned}
(d_{i11} - d_{i21}) &= (\alpha_{11} - \alpha_{21}) + (\epsilon_{i11} - \epsilon_{i21}),\\
(d_{i11} - d_{i22}) &= (\alpha_{11} - \alpha_{22}) + (d^*_{i1} - d^*_{i2}) + (\epsilon_{i11} - \epsilon_{i22}),\\
(d_{i21} - d_{i22}) &= (\alpha_{21} - \alpha_{22}) + (d^*_{i1} - d^*_{i2}) + (\epsilon_{i21} - \epsilon_{i22}).
\end{aligned} \tag{2.15}
$$

The equations in (2.15) have five unknowns, and therefore, cannot be used to identify the individual additive errors without some restrictions. Taking the average across sections i of the differences in (2.15) allows the identification of the differences between the additive errors of the technologies used. This is demonstrated below:

The first equation in (2.15) gives:

$$(\overline{d_{11}} - \overline{d_{21}}) = (\alpha_{11} - \alpha_{21}),$$

which is an estimate of the differences between additive errors of technology $j = 1$ and $j = 2$, when measuring distress type $k = 1$.

And the second and third equations give:

$$
\begin{aligned}
(\overline{d_{11}} - \overline{d_{22}}) &= (\alpha_{11} - \alpha_{22}) + (\overline{d^*_1} - \overline{d^*_2}),\\
(\overline{d_{21}} - \overline{d_{22}}) &= (\alpha_{21} - \alpha_{22}) + (\overline{d^*_1} - \overline{d^*_2}),
\end{aligned}
$$

which cannot be used to identify the biases when the same technology is measuring different distress types without further assumptions. For example, one can assume that $\alpha_{22} = \alpha_{21}$, that is, additive biases are only technology specific, in which case we can identify the difference between the true distresses of type $k = 1$ and $k = 2$, from the third equation.

Let

$$\bar{d}_{ij} = \frac{1}{K}\sum_{k=1}^{K} d_{ijk}.$$

Then:

$$\overline{d_{ij}} = \overline{\alpha_j} + \overline{d^*_i} + \overline{\epsilon_{ij}}, \tag{2.16}$$

where:

$$
\begin{aligned}
\overline{\alpha_j} &= \frac{1}{K}\sum_{k=1}^{K} \alpha_{jk};\\
\overline{d^*_i} &= \frac{1}{K}\sum_{k=1}^{K} d^*_{ik};\\
&\text{and}\\
\overline{\epsilon_{ij}} &= \frac{1}{K}\sum_{k=1}^{K} \epsilon_{ijk}.
\end{aligned} \tag{2.17}
$$

Equations (2.16) and (2.17) can be used to estimate the overall biases of the technologies $j = 1, \ldots, J$ in measuring K items by using differencing techniques as demonstrated in (2.15). The individual additive errors, however, cannot be identified without further assumptions, such as those discussed above.

Although this specification allows the inclusion of additive systematic biases and random errors of measurement, the values of the biases cannot be independently identified. The differences between the observed measurements by different technologies can only be used to estimate the differences between the biases of technologies measuring the same items, as demonstrated above.

Because there are strong reasons to expect multiplicative biases in measurements of condition for the case of infrastructure inspection, only those techniques that allow inclusion of multiplicative errors will be considered. Therefore, this type of model will not be considered further.

The other type of models commonly used are experimental and quasi-experimental techniques that are discussed next.

Estimation alternatives: Experimental and quasi-experimental techniques

A variety of experimental and quasi-experimental techniques have been used to estimate the size of different types of measurement errors. The problem can be stated as follows: Consider the generalized measurement equation in (2.2). Let the deviation between the true value d^*_{ik} and the measured value d_{ijk} be ϵ_{ijk}'. Then the measured value can be related as:

$$d_{ijk} = d^*_{ik} + \epsilon_{ijk}', \tag{2.18}$$

where ϵ_{ijk}' is the total error of measurement due to the jth technology, measuring item k, in location i.

There are three sources of variation in eqn (2.18), location specific, technology specific and item specific. The models presented in the previous sections made certain assumptions about these sources of error to allow for the estimation of the measurement bias parameters. Experimental design techniques control these sources of variation. There are a variety of statistical models for three sources of variation that are commonly used in experimental design. Since the models discussed in the past section have been linear, first-order models, the following discussion will concentrate on this type of model.

The first-order experimental design model assumes that the effects of influencing factors on the results of measurement (response) are linear and additive. Let the additive error term in eqn (2.2) be defined as follows:

$$\alpha_{jk} = \alpha_0 + \theta_j + \theta_k + \theta_{jk}, \tag{2.19}$$

where:

α_0 = overall additive error,
θ_j = technology specific effect,
θ_k = item (distress) specific effect, and
θ_{jk} = interactions between technology and distress effects.

The model in (2.19) assumes that, the additive error of measuring distress type k by technology j can be separated into its main linear effects that are specific to the technology and distress types. The interactions between technology and distress specific effects and other higher-order interactions can be assumed negligible in eqn (2.19) for convenience.

Substituting eqn (2.19) in eqn (2.2) without interactions gives:

$$d_{ijk} = \alpha_0 + \beta_{jk} d^*_{ik} + \theta_j + \theta_k + \epsilon_{ijk}. \tag{2.20}$$

Assuming there is no multiplicative bias, $\beta_{jk} = 1$, eqn (2.20) becomes:

$$d_{ijk} = d^*_{ik} + \alpha_0 + \theta_j + \theta_k + \epsilon_{ijk}, \tag{2.21}$$

which is the widely used linear model in experimental design (Cochran and Cox, 1950). Taking the overall mean, across all items (distresses), locations (sections), and observations (technologies) of eqn (2.21) gives:

$$\overline{d} = \alpha_0 + \overline{\theta} + \overline{d^*} + \overline{\epsilon},$$

where $\overline{d}$ = overall mean across all technologies J, items K, and locations I. Assuming $\overline{\theta} = 0$, and $\overline{\epsilon} = 0$, then:

$$\alpha_0 = \overline{d} - \overline{d^*}.$$

Substituting this result in eqn (2.21) gives:

$$d_{ijk} = (\overline{d} - \overline{d^*}) + \theta_j + \theta_k + d^*_{ik} + \epsilon_{ijk},$$

$$d_{ijk} = \overline{d} + \theta_j + \theta_k + \epsilon'_{ijk}, \tag{2.22}$$

where:

$$\epsilon'_{ijk} = (d^*_{ik} - \overline{d^*}) + \epsilon_{ijk}.$$

Assumptions made in estimating this model are:

$$\sum_{j=1}^{J} \theta_j = 0,$$
$$\sum_{k=1}^{K} \theta_k = 0. \tag{2.23}$$

The model in eqn (2.22) is similar to the model in eqn (2.6), where additive biases are accounted for. Additionally, eqn (2.22) allows the inclusion of the effects of distress factors affecting the results of measurement captured by the term θ_k. The assumptions in (2.23) are valid if the items (distresses) are encountered randomly by the technologies j, which is the case in infrastructure inspection.

The analysis of variance approach is able to detect the presence of measurement errors and distribute them among the various sources. The approach, however, suffers from similar limitations to those discussed in previous sections. Specifically, the assumption that the average across all items and technologies $\overline{d}$ represents the true average. This is demonstrated below. Recall the relationship in eqn (2.2).

Assuming that:

$$\overline{\alpha} = 0,$$
$$\overline{\beta} = 1.$$

Taking the average across all locations, items and technologies of eqn (2.2) gives:

$$\overline{d} = \overline{\alpha} + \overline{\beta d^*} + \overline{\epsilon}. \tag{2.24}$$

Substituting (2.24) in (2.22) gives:

$$d_{ijk} = \overline{\alpha} + \overline{\beta d^*} + \theta_j + \theta_k + \epsilon'_{ijk} + \overline{\epsilon}.$$

Then:

$$d_{ijk} = \overline{d^*} + \theta_j + \theta_k + \epsilon_{ijk} + \overline{\epsilon}.$$

Assuming $\overline{\epsilon} = 0$, this becomes the model in (2.22).

However, if $\overline{\epsilon}$ is not zero, then the estimation of (2.22) is biased. This is likely, as the models in (2.22) assume there is no multiplicative bias, and there is reason to believe that there is substantial multiplicative bias in some technologies.

The advantage of experimental design techniques is the ability to control the influence of extraneous factors of variation. For example, to be able to justify the model with additive bias in eqn (2.6), one can design a measurement process that "averages" out the effects of the location and item-specific factors. There are several types of designs that are typically used for this, such as randomized block designs and Latin square designs.

Since the assumptions made in the literature were found to be either too restrictive, as in the assumption on multiplicative bias, or lead to erroneous estimates, as with the assumption about an unbiased technology used as a reference, there is a need to develop an alternate methodology to study measurement errors. A proposed alternative is the use of factor analytic modeling techniques, which do not require knowledge of the true value of an item, to formulate the measurement problem. The proposed formulation is presented next. Estimation procedures and data collection needs for such an approach can be found in Humplick (1989).

3. THE FACTOR ANALYTIC APPROACH

This section presents the factor analytic specification of a measurement error model. Defining the measurement error model in matrix notation for each location i and measurement process j type gives:

$$\underset{(I\times J)}{D} = \underset{(I\times 1)}{I_I}\ \underset{(1\times J)}{\alpha} + \underset{(I\times 1)}{D^*}\ \underset{(1\times J)}{\beta} + \underset{(I\times J)}{\epsilon}, \tag{3.1}$$

where:

D = $(I \times J)$ matrix of measured observations,
I_I = $(I \times 1)$ vector of ones,
α = $(1 \times J)$ vector of additive systematic measurement errors that are specific to a technology j,
β = $(1 \times J)$ vector of multiplicative systematic measurement errors that are specific to a technology j,
D^* $(I \times 1)$ vector of unknown true values,
ϵ = $(I \times J)$ matrix of random measurement errors that are technology and item related.

The parameters α and β are independent of location characteristics as the locations are assumed randomly selected from a population of locations with varying measurement characteristics. The specification in (3.1), therefore, represents both technology and item (distress) specific biases.

Each element in (3.1) can be expressed as:

$$d_{ij} = \alpha_j + \beta_j d_i^* + \epsilon_{ij}. \tag{3.2}$$

The objective is to estimate the technological biases α_j, β_j and the variance of measurement Var (ϵ_{ij}) for each location type.

The type of methods discussed previously apply calibration techniques such as least squares regression to estimate measurement errors, and are termed the "regression approach". The second type of methods apply covariance analysis techniques to estimate measurement errors, and are termed the "factor-analytic approach". The methodology developed in this paper uses the factor-analytic approach.

The form of the factor analysis model which is used in this paper, is the single factor model. This is a variation of what is commonly known as the "common factor model", where all the variables are related to a single common factor (Rummel, 1970). The single factor model can be derived from the measurement error equation in (3.1) as follows:

Recall the observable random matrix **D** of distress measurements with $I \times J$ components, and elements d_{ij}, where I is the number of sections and J is the number of technologies used. These technologies can be thought of as representing different indicators of the true distress on a section. Let the mean of this matrix taken across I be $\overline{D}$, with $1 \times J$ components and elements $\bar{d}_j$. The true distress is represented by the random vector D^*

in eqn (3.1), with $I \times 1$ elements d_i^*, which is unobserved. This vector is random, if the sections inspected are randomly selected from a population of pavement segments with varying types and levels of distress as discussed earlier. Let the mean of the vector D^* be the scalar $\overline{d^*}$. The common factor model postulates that **D** is linearly related to the vector of latent random variables D^* with $I \times 1$ components called the *common factor*, and to $I \times J$ additional sources of variation called *specific variances* or *errors*. The errors are represented by the variance of the random matrix ϵ, with $I \times J$ components and elements ϵ_{ij}, defined as ψ with $(J \times J)$ elements ψ_{jl}, where j and l are two technologies.

The average across all observations (sections I of eqn (3.1) gives:

$$\underset{(1\times J)}{\overline{D}} = \underset{(1\times J)}{\alpha} + \overline{d^*} \underset{(1\times J)}{\beta} + \underset{(1\times J)}{\bar{\epsilon}}, \tag{3.3}$$

where:

$\bar{\epsilon}$ with $(1 \times J)$ components is the average random error specific to a technology j. Or for each element in the above expression:

$$\overline{d_j} = \alpha_j + \beta_j\overline{d^*} + \overline{\epsilon_j} \qquad \forall j = 1, \ldots, J. \tag{3.4}$$

Premultiplying eqn (3.3) by l_I and subtracting it from (3.1) gives the following:

$$\underset{(I\times J)}{D - l_I\overline{D}} = \underset{(I\times 1)}{(D^* - l_I\overline{d^*}))} \underset{(1\times J)}{\beta} + \underset{(I\times J)}{\epsilon} - \underset{(I\times J)}{l_I\,\bar{\epsilon}}. \tag{3.5}$$

Or for each element in the above expression:

$$(d_{ij} - \overline{d_j} = \beta_j(d_i^* - \overline{d^*}) + (\epsilon_{ij} - \overline{\epsilon_j}). \tag{3.6}$$

The coefficient β_j is called the *loading* of the jth variable on the common factor $(d_i^* - \overline{d^*})$. The $I \times J$ deviations $(d_{ij} - \overline{d_j})$ are expressed in terms of I random variables $(d_i^* - \overline{d^*})$ and $I \times J$ error terms $(\epsilon_{ij} - \overline{\epsilon_j})$ which are latent.

The deviation $(d_{ij} - \overline{d_j})$ of the observed measurement on section i by technology j from the mean-observed distress by j taken across all sections i is expressed in terms of the unobserved deviation $(d_i^* - \overline{d^*})$ of the true distress on the section i from the mean true distress taken across all sections i, and the random error of measurement. The coefficient β_j represents how much of the true distress is captured by technology j, that is, how heavily technology j loads on the common factor $(d_i^* - \overline{d^*})$. Thus, if multiple technologies $j = 1, \ldots, J$ are used, the contribution of each technology to the measurement of the true distress on a section is measured by this coefficient. Therefore, β_j measures the amount of true distress picked up by technology j, which was defined as the multiplicative bias. If $\beta_j = 1$, then technology j has no multiplicative bias. If $\beta_j < 1$, then technology j systematically underestimates the true distress by the amount $1 - \beta_j$, for the case where there is no additive bias ($\alpha = 0$). Similarly, if there is no additive bias, and $\beta_j > 1$, then technology j overestimates the true distress by the amount $\beta_j - 1$.

The estimation procedure to obtain the parameters in eqn (3.1) using the factor analytic approach is discussed next.

Let the population covariance between the measurements by the J technologies be Σ with $J \times J$ components and elements θ_{jl}, where j and l are two technologies. Let the sample covariance matrix S, with elements S_{jl}, be an estimator of the unknown population covariance matrix Σ. The estimation of the parameters in (3.1) from the covariance matrix S is now outlined. A factor analysis estimation method is used to estimate the systematic multiplicative bias and the random error of measurement specific to a technology j (β_j,var($\epsilon_{ij} - \overline{\epsilon_j}$ $= \psi_j$,) and the true value of distress $(d_i^* - \overline{d^*})$. Then, a normalization assumption is specified to identify d_i^* and α_j. The step-by-step process is outlined below.

Estimation of the multiplicative bias

The multiplicative bias β_j and the precision of measurement $\mathrm{var}(\epsilon_{ij} - \bar{\epsilon}_j) = \psi_j$ are estimated by the method of unweighted least squares (ULS). The method uses the following logic.

The covariance between the measurements by the various technologies J has the following structure:

$$\begin{aligned}
\underset{(J\times J)}{\Sigma} &= \mathrm{Cov}(D) = E\left[\underset{(J\times I)}{(D - l_I\bar{D})'}\underset{I\times J)}{(D - l_I\bar{D})}\right] \\
&= E[[(D^* - l_I\overline{d^*})\beta + (\epsilon - l_I\bar{\epsilon})]'[(D^* - l_I\overline{d^*})\beta + (\epsilon - l_I\bar{\epsilon})]] \\
&= E[[\beta'(D^* - l_I\overline{d^*})' + (\epsilon - l_I\bar{\epsilon})'][(D^* - l_I\overline{d^*})\beta + (\epsilon - l_I\bar{\epsilon})]] \\
&= \underset{(J\times 1)}{\beta'}\, E\left[\underset{(1\times I)}{(D^* - l_I\overline{d^*})'}\underset{(I\times 1)}{(D^* - l_I\overline{d^*})}\right]\underset{(1\times J)}{\beta} + E\left[\underset{(J\times I)}{(\epsilon - l_I\bar{\epsilon})'}\underset{(I\times J)}{(\epsilon - l_I\bar{\epsilon})}\right] \qquad (3.7) \\
&\quad + 2E\left[\underset{(J\times I)}{(\epsilon - l_I\bar{\epsilon})'}\underset{(I\times 1)}{(D^* - l_I\overline{d^*})}\underset{(1\times J)}{\beta}\right], \\
\Sigma &= \beta'\beta\,\mathrm{Var}(D^*) + \mathrm{Cov}(\epsilon) + 2\,\mathrm{Cov}(\epsilon,D^*)\beta.
\end{aligned}$$

The covariance between the observed matrix D and the latent vector D^* has the following structure:

$$\begin{aligned}
\underset{(J\times 1)}{\mathrm{Cov}(D,D^*)} &= E\left[\underset{(J\times I)}{(D - l_I\bar{D})'}\underset{(I\times 1)}{(D^* - l_I\overline{d^*})}\right] \\
&= E[[(D^* - l_I\overline{d^*})\beta + (\epsilon - l_I\bar{\epsilon})]'[D^* - l_I\overline{d^*}]] \\
&= E\left[\underset{(J\times 1)}{\beta'}\underset{(1\times I)}{(D^* - l_I\overline{d^*})'}\underset{(I\times 1)}{(D^* - l_I\overline{d^*})} + \underset{(J\times I)}{(\epsilon - l_I\bar{\epsilon})'}\underset{(I\times 1)}{(D^* - l_I\overline{d^*})}\right] \qquad (3.8) \\
\mathrm{Cov}(D,D^*) &= \beta'\,\mathrm{Var}(D^*) + \mathrm{Cov}(\epsilon,D^*).
\end{aligned}$$

A number of assumptions are required to estimate β and $\mathrm{Cov}(\epsilon) = \psi$ from eqns (3.7) and (3.8). These are summarized below:

(a) The variability of the true value of distress across sections is unknown and has a variance denoted by σ^2, which depends on how the sections have been selected.

$$\underset{(1\times 1)}{\mathrm{Var}(D^*)} = E\left[\underset{(1\times I)}{(D^* - l_I\overline{d^*})'}\underset{(I\times 1)}{(D^* - l_I\overline{d^*})}\right] = \underset{(1\times 1)}{\sigma^2}.$$

It is assumed that the sections I are randomly selected, and do not come from a single stretch of pavement.

(b) The random error of distress measurement on a section is independent of the random error of distress measurement on other sections, and the variance of the random error is specific to a given technology. This will be referred to as the *specific variance* of a technology.

$$\underset{(J\times J)}{\mathrm{Cov}(\epsilon)} = E\left[\underset{(J\times I)}{(\epsilon - l_I\bar{\epsilon})'}\underset{(I\times J)}{(\epsilon - l_I\bar{\epsilon})}\right] = \underset{(J\times J)}{\Psi}.$$

It is assumed that Ψ is a diagonal matrix. This assumption is justified on the basis that the technologies employed are radically different, and their measurement errors are not correlated because they employ different measuring principles and procedures. That is, the factors causing errors to one technology do not affect another technology in the same manner, although they are measuring the same sections. If the technologies are measuring distresses in sections with homogeneous section characteristics, such as flexible pavements, this assumption is justified. If similar technologies are used, such as two photographic techniques measuring the same sections, this assumption may not be valid. There are techniques that allow specification of the correlation between the random errors of measurement by different technologies. These were not employed in this paper as the model specified in (3.4) is estimated independently for a group of sections of the same pavement type, and hence does not include section characteristics.

(c) The true value of distress on a section D^* is independent of the random error of measurement ϵ for a given technology in that section.

$$\underset{(J\times 1)}{\mathrm{Cov}(\epsilon,D^*)} = E\left[\underset{(J\times I)}{(\epsilon - l_I\bar{\epsilon})'}\underset{(I\times 1)}{(D^* - l_I\bar{d}^*)}\right] = \underset{(J\times 1)}{0} .$$

This assumption is justified on the basis that the effects of the true value of distress d_i^* on the results of measurement are captured by β_j.

Using these assumptions and eqns (3.7) and (3.8) yields:

$$\Sigma = \beta'\beta\sigma^2 + \psi, \tag{3.9}$$

$$\mathrm{Cov}(D,D^*) = \beta'\sigma^2. \tag{3.10}$$

The relationship in (3.10) is not observed but it is part of the covariance structure implied by the factor-analytic model specification. We have three types of unknowns β, σ, ψ in eqn (3.9). Here β has J elements, σ is a scalar, and ψ has J elements, a total of $(2J + 1)$ unknowns. On the other hand, we observe $(J + J(J - 1)/2) = J(J + 1)/2$ covariances S. The relationship in (3.9) is therefore over-identified as long as $J \geq 4$. For example, if D contains $J = 7$ technologies, and the model in (3.2) is appropriate, the $J(J + 1)/2 = 7(8)/2 = 28$ elements of Σ are described in terms of $2J + 1 = 15$ parameters β_j,ψ_j and σ of the factor-analytic model in (3.9). Therefore, the necessary conditions for identification are met.

Equation (3.9), however, can be rewritten as follows:

$$\Sigma = (\beta'\sigma)(\beta\sigma) + \psi. \tag{3.11}$$

Thus, the vector of parameters β is multiplied by σ and therefore it is not possible to identify both β and σ. The restriction most commonly used is to set the variance of the latent variable σ^2, to one. This restriction has no theoretical justification, but most factor-analytic software use this restriction. Alternatively one can set one of the multiplicative error terms, say β_r for a reference technology r, to one. This restriction has theoretical justification, since $\beta_r = 1$, if the technology r has no multiplicative bias. The estimation software that was available did not allow the latter restriction. However, as is demonstrated below, the first restriction can be used, and the second restriction is derived from the results of estimation using the first restriction. Thus the two restrictions are equivalent.

Each element in the matrices in (3.11) will have the following relationship:

$$\begin{aligned}
\sigma_{jl} &= (\beta_j\sigma)(\beta_l\sigma) + \psi_{jl} \qquad \forall j \neq l \quad (\psi_{jl} = 0), \\
\sigma_{jj} &= (\beta_j\sigma)(\beta_j\sigma) + \psi_j = \beta_j^2\sigma^2 + \psi_j \qquad \forall j, \\
\beta_j^2 &= (\sigma_{jj} - \psi_j)/\sigma^2.
\end{aligned} \tag{3.12}$$

A reference technology r will have the following relationship:

$$\begin{aligned}
\sigma_{rr} &= \beta_r^2\sigma^2 + \psi_r, \\
\beta_r^2 &= (\sigma_{rr} - \psi_r)/\sigma^2.
\end{aligned} \tag{3.13}$$

Dividing (3.13) by (3.14) gives:

$$\frac{\beta_j^2}{\beta_r^2} = \frac{\sigma_{jj} - \psi_j}{\sigma_{rr} - \psi_r}. \tag{3.14}$$

The result in (3.14) shows that, if we set $\sigma^2 = 1$, and estimate the β's, then divide the result by a β_r that we select as a reference, the estimation results are equivalent to those obtained by setting $\beta_r = 1$. This approach is used to estimate the multiplicative biases using software that sets $\sigma^2 = 1$.

An expression for the additive error α_j can be obtained by rearranging eqn (3.4)

$$\alpha_j = \overline{d_j} - \beta_j\overline{d^*} - \overline{\epsilon}_j. \tag{3.15}$$

This requires knowledge about $\overline{d^*}$, which has not been discussed so far in the specification of the factor-analytic model. The type of assumptions required to identify α_j and $\overline{d^*}$ are discussed in the next section, and are termed "normalization restrictions".

The estimation procedure to obtain α_j, β_j, d_i^*, and $\mathrm{var}(\epsilon_{ij}) = \psi_j$ from eqns (3.12) and (3.15) is outlined next.

If the sample covariance matrix is S, unweighted least squares seeks an $\hat{S}$ such that:

$$\hat{S} = (\hat{\beta}'\sigma)(\hat{\beta}\sigma) + \hat{\psi}. \tag{3.16}$$

The principle behind the ULS approach is to minimize the residual covariance Q where:

$$Q = \Sigma_j\Sigma_l(s_{jl} - \hat{s}_{jl})^2, \tag{3.17}$$

so that the Q obtained is smaller than any other quantity we can get by choosing other values of $(\hat{\beta}\sigma)$ and $\hat{\psi}$ in equation (3.16). This gives the best estimates of $(\hat{\beta}_j\sigma)$ and $\hat{\psi}_j$. If we select a technology r which is a reference, and divide $(\beta_j\sigma)$ by $(\hat{\beta}_r\sigma)$ we obtain $\hat{\beta}_j$. The starting values of $\hat{s}_{jl}$ most commonly used are one, or the squared multiple covariance of a variable j with all other variables $j = 1, \ldots, (J - 1)$. The squared multiple covariance is the percentage variance which the variable j has in common with all the other variables in a data matrix.

4. COMPARISON OF RESULTS

The models described in Section 2 and 3 were estimated from the results of a Federal Highway Administration (FHWA) study on "Improved Methods and Equipment to Conduct Pavement Distress Surveys" (Hudson, Elkins, Uddin and Reilley, 1987). The study included measurement of distresses on experimental units with varying densities of distress occurrence and contrast between distresses and the background on which they appear.

The technologies included three direct measurement technologies (mapping, manual, logging) involving visual inspection by humans; and four indirect measurement techniques involving optic (photo1, photo2, video) and laser technologies. The technologies had a wide range of capabilities in terms of the resolution of measurement, sampling size and data processing and reduction. All seven technologies were used to measure distresses on all experimental units, but due to incommensurate units, some of the results were not usable. The data set consisted of 490 observations.

Four alternate models were estimated based on these data. These were: (1) factor analysis normalized to a reference technology (FAREF); (2) factor analysis normalized to a constructed average (FAAV); (3) ordinary least squares regression against a reference technology (OLSREF); and (4) ordinary least squares regression against a constructed average (OLSAV). These models were estimated independently for each distress type (out of seven measured distresses). However, due to the large number of estimations (7 × 7

$\times$ 4 = 196), only one distress type (alligator cracking) will be used to illustrate the results. Alligator cracking is an indicator of structural distress on flexible (asphalt) pavements, which has a scaly appearance like alligator skin. The results of the entire study can be found in Humplick (1989).

Overall model adequacy

The estimation results for the four models are compared in this section to demonstrate the superiority of the factor-analytic specifications. The estimation procedures used are documented in Ben-Akiva and Humplick (1990).

There are three main elements of the estimation results that can be compared:

1. Vectors of multiplicative biases ($\hat{\beta}_j$), which represent the estimated fraction of true distress picked up by technology j.
2. Vectors of the standard deviation of the random term S.D.(ϵ_{ij}) = ($\sqrt{\Psi_j}$), which represents the variance of measurement specific to a technology j.
3. Vectors of the coefficient of determination $(R_j)^2$ for each technology j, which measures the overall model fit to the data.

Table 1 compares factor analysis normalized to a reference (FAREF) with regression against the same reference (OLSREF). The reference technology used was manual inspection, and it was assumed that this reference was unbiased. There were no results for the laser technology, as it measured alligator cracking in units that could not be converted to those of the other technologies. However, measurements for other distress types using this technology were included in the estimation, and appear in Humplick (1989).

The estimated multiplicative biases by regression are 5 to 10% smaller than those by factor analysis. This demonstrates the expected estimation bias of regression from the derivations in Section 2. The pattern of the biases, however, is the same for both methodologies, with photo2 having the highest and video having the lowest estimated multiplicative bias. Shifts in the pattern of multiplicative biases between the two methodologies would be an indication that there are underlying differences in the way the methodologies relate to the data.

The estimated standard deviations of the random term were larger in the case of the regression model (OLSREF) as compared to the factor analytic specification (FAREF). This indicates that the factor-analytic specification is capturing more of the variation in the data in the parameters (α_j,β_j), than the regression model.

Finally, there is an improvement in the coefficient of determination for the factor analysis model (FAREF) over that for the regression model (OLSREF). This indicates that the factor-analytic specification fits the data better.

Table 1. Factor analysis vs. OLS regression against a reference technology

Case of alligator cracking on flexible pavements
Estimated parameters (standard errors of the estimates)

Technology (j)	$\hat{\alpha}_j$ (sq ft) OLSREF	$\hat{\alpha}_j$ (sq ft) FAREF	$\hat{\beta}_j$ OLSREF	$\hat{\beta}_j$ FAREF	S.D. (ϵ_{ij}) = $\sqrt{\Psi_j}$ (sq ft) OLSREF	S.D. (ϵ_{ij}) = $\sqrt{\Psi_j}$ (sq ft) FAREF	coeff. of det. R_j^2 OLSREF	coeff. of det. R_j^2 FAREF
Mapping	−76.9	−140.2	1.59	1.70	529.2	396.9	0.89	0.94
	(221.2)	(275.5)	(0.20)	(0.21)				
Manual	0	0	1.00	1.00	–	262.9	–	0.94
Logging	589.7	466.9	2.43	2.64	2064.0	646.2	0.56	0.61
	(862.2)	(530.8)	(0.76)	(0.78)				
Photo1	−186.6	−253.3	2.14	2.25	538.3	444.5	0.93	0.95
	(225.4)	(262.1)	(0.20)	(0.19)				
Photo2	−474.6	−657.6	3.49	3.80	1187.4	551.3	0.99	0.99
	(495.9)	(407.7)	(0.44)	(0.46)				
Video	135.8	99.1	0.84	0.90	489.9	472.3	0.73	0.77
	(203.7)	(262.1)	(0.18)	(0.19)				
Laser	–	–	–	–	–	–	–	–

Table 2. Factor analysis vs. OLS regression against an average

Case of alligator cracking on flexible pavements
Estimated parameters (standard errors of the estimates)

Technology (j)	$\hat{\alpha}_j$ (sq ft) OLSAV	$\hat{\alpha}_j$ (sq ft) FAAV	$\hat{\beta}_j$ OLSAV	$\hat{\beta}_j$ FAAV	S.D.(ϵ_{ij}) = $\sqrt{\Psi_j}$ (sq ft) OLSAV	S.D.(ϵ_{ij}) = $\sqrt{\Psi_j}$ (sq ft) FAAV	coeff. of det. R_j^2 OLSAV	coeff. of det. R_j^2 FAAV
Mapping	−46.8	−73.0	0.81	0.83	458.3	396.9	0.92	0.94
	(189.2)	(474.2)	(0.08)	(0.17)				
Manual	52.7	37.5	0.48	0.49	714.1	262.9	0.91	0.94
	(115.4)	(363.7)	(0.05)	(0.10)				
Logging	433.3	570.0	1.41	1.29	1546.0	646.2	0.75	0.61
	(638.1)	(845.1)	(0.29)	(0.54)				
Photo1	−123.1	−154.5	1.06	1.09	565.7	444.5	0.93	0.95
	(234.4)	(527.0)	(0.11)	(0.21)				
Photo2	−470.1	−501.3	1.82	1.85	608.3	551.3	0.97	0.99
	(250.6)	(551.6)	(0.11)	(0.23)				
Video	154.0	134.0	0.42	0.44	812.4	472.3	0.74	0.77
	(195.5)	(474.2)	(0.09)	(0.17)				
Laser	–	–	–	–	–	–	–	–

Therefore, the factor analysis specification normalized to a reference technology resulted in better fit than the OLS regression against the same reference. This demonstrates that the factor-analytic specification, which assumes the true value of distress is latent, is a better calibration procedure, if multiple measurements are available. The advantage of regression, however, is that one needs a minimum of two technologies to run the calibration, as opposed to a minimum of three technologies in the case of factor-analytic specifications. The results of calibration procedures are used for future decision making such as: planning maintenance, predicting deterioration, and improving inspection technologies. An accurate calibration procedure is, therefore, necessary.

A similar comparison between the factor analysis model normalized to a constructed average (FAAV) and regression against the same average (OLSAV) is shown in Table 2. The average was constructed over a combination of measurements from six radically different technologies, assuming that the biases would cancel out. Again, the measures by the laser technology were not included.

The difference between the estimated results were not as large as in Table 1. The estimated multiplicative biases were still lower for the case of regression, supporting the derivation in Section 2. The superiority of factor analysis, however, is not as clear as in the case of Table 1. This is due to the averaging, which smooths out a lot of the noise before regression, as shown below:

$$d_{ij} = \alpha_j + \beta_j \frac{1}{T} \sum_{j=1}^{T} d_{ij} + \epsilon_{ij}, \tag{4.1}$$

where T is the number of technologies used to construct the average. Since d_{ij} appears also on the right-hand side, the standard errors of the regression are smaller, and hence, the fit much better.

Informal test of the parameter estimates

An informal test comparing analogous values from the two model specifications (factor analytic and regression) was developed. This involved the comparison of the Euclidian distance between vectors of estimated parameters from the different model specifications.

Let (OLS) denote the regression model and (FA) the factor- analytic model. Define the vectors of estimated parameters for a technology j by η_j^{OLS} and η_j^{FA}, respectively, where η can be any one of the elements discussed earlier (β_j, $\sqrt{(\psi_j)}$, R_j^2). Then the Euclidian distance between these vectors is defined by:

$$\mu_{OLS,FA} = \left[\sum_{j=1}^{P}\left(\frac{\eta_j^{OLS} - \eta_j^{FA}}{P}\right)^2\right]^{1/2}, \tag{4.2}$$

where:

$\mu_{OLS,FA}$ = root mean square difference between model OLS and model FA,
η_j^{OLS} = estimated vector of parameters for technology j by model OLS,
η_j^{FA} = estimated vector of parameters for technology j by model FA,
P = number of technologies common to both models.

This test gives a measure of the distance between the parameter estimates. Since the above test is informal, it cannot be used to make statistical statements about models OLS and FA. A formal test that allows such statements is developed in Humplick (1989).

The results of the informal test on the equality of estimated parameters $(\alpha_j, \beta_j, \psi_j)$, from different model specifications are shown in Table 3. Models FAAV and OLSAV display estimates of multiplicative bias that are closer in agreement with ($\mu_{OLS,FA} = 0.05$) than models FAREF and OLSREF with ($\mu_{OLS,FA} = 0.17$). Similarly for the additive error with $\mu_{OLS,FA} = 60.50$ and $\mu_{OLS,FA} = 90.63$; and the variance of the random error (precision) with $\mu_{OLS,FA} = 1.00$ and $\mu_{OLS,FA} = 1.78$, respectively for the comparisons between the four models. This test is an indication of the severity of the violations in the regression against a reference (OLSREF) as compared to those of the regression against a constructed average (OLSAV).

5. CONCLUSIONS

The theory and methods of analyzing measurement errors were developed in this paper. This included a review of state-of-the-art techniques in measurement error modeling. Measurement error theory was applied to the highway pavement surface condition inspection problem, where a methodology for estimating measurement errors using factor-analytic specifications was developed. Data from highway pavement surface condition inspections was used to test the methodology developed against the regression approach.

The factor-analytic and regression specifications both require the use of a reference measurement for calibration. Two reference measurements were used: (i) assuming there is an unbiased measurement technique; and (ii) assuming the average over radically different technologies is unbiased. This resulted in four model specifications: (i) factor analysis normalized to a reference (FAREF); (ii) factor analysis normalized to an average (FAAV); (iii) ordinary least squares regression against a reference (OLSREF); and (iv) ordinary least squares against an average (OLSAV).

The factor-analytic specifications were superior to regression in terms of giving better fit to the data, and resulting in unbiased parameters. Additionally, the assumptions in the specification and estimation of the factor-analytic models were supported theoretically, from the expected behavior of measurement errors. The regression models required assumptions that were easily violated when applied to the data set used. For example, the assumption that the true distress can be replaced by a measured distress from an unbiased technology, resulted in biased estimates of the measurement errors. The size of the bias due to regression was about 5 to 10% as compared to the factor analytic estimates.

Table 3. Informal test on the assumptions of different model specifications. Case of alligator cracking on flexible pavements

Parameter (η_j)	Root mean square differences (μ) OLSREF vs. FAREF	OLSAV vs. FAAV
Additive error α_j	98.63	60.50
Multiplicative error β_j	0.17	0.05
Random error Ψ_j	1.78	1.00

Although regression is widely used as a calibration technique in other disciplines, it falls short when applied to infrastructure inspection situations. This is mainly due to the lack of well developed and precise measurement technologies whose measurements can be assumed to replace the true distress.

The size of the bias in estimates of measurement errors by regression is greatly reduced by simple averaging. This is demonstrated by the closer agreement between the results of factor analysis normalized to an average when compared to regression against an average, as opposed to factor analysis normalized to a reference technology compared to regression against a reference technology.

The factor-analytic approach required two types of assumptions for estimation and specification. The first type was assumptions about the data and relationships between the main variables in the data. These were assumptions such as the nature of the distribution of true distress in the data set used, the relationship between the random errors of measurement from observation to observation, and the relationship between the random errors of measurement and the true distress. These assumptions were justified on the basis that the distress data were collected from random sections, and the models were estimated on observations with homogeneous section (location) characteristics. These assumptions allowed the specification of the factor-analytic model, and were referred to as "specification assumptions". The second type was assumptions about the nature of the observations from a certain group of variables. These were assumptions such as the availability of a reference measurement which came from a set of unbiased measurements by a reference technology, or from an unbiased estimate of true distress obtained from various groupings of observations (e.g. an average from a set of technologies). These assumptions allowed the estimation of the factor-analytic model, and were termed "normalization restrictions". The normalization restrictions had to be justified extraneously from knowledge about the technologies of measurement.

The choice of a reference measurement to use for normalization is not a straightforward procedure, and a combination of subjective knowledge about the technologies and estimation results is required. A systematic approach to selecting a reference measurement is treated in Humplick (1989).

The methodology developed can be applied to a wide range of problems. It can be used to calibrate existing technologies and develop correction factors to adjust inspection and measurement results. It can also be used to choose among a set of existing technologies, or to decide on the measurement situation for which a particular technology is best suited.

Acknowledgements—This work is part of a Ph.D. dissertation in the Department of Civil Engineering at MIT under the supervision of Moshe Ben-Akiva. Support for the research was obtained from the U.S. Army Research Office through funding for the Center for Advanced Construction Technology in the Center for Construction Research and Education at MIT. Thanks are due to Moshe Ben-Akiva, Sue McNeil, and Frank Koppelman for thoughtful reviews, and Samer Madanat, Rabi Mishalani, and anonymous reviewers for editorial assistance. The views, opinions and findings contained in this paper are those of the author.

REFERENCES

Ben-Akiva M. and Humplick F. (1990) A methodology for estimating the accuracy of inspection systems. Working Paper presented at the October 1990 ORSA, Special session on Infrastructure, Philadelphia.

Cochran W. G. and Cox G. M. (1950) *Expermental Designs*. Wiley, New York.

Cochran W. G. (1968) Errors of measurement in statistics. *Technometrics*, **10**(4), 637–666.

Fellegi, I. (1964) Response variance and its estimation. *J. Amer. Stat. Assoc.*, **59**, 1016–1041.

Grubbs F. E. (1948) On estimating precision of measuring instruments and product variability. *J. Amer. Stat. Assoc.*, **43**, 243–264.

Grubbs F. E. (1973) Errors of measurement; Precision, accuracy, and the statistical comparison of measuring instruments. *Technometrics*, **15**(1), 53–66.

Hahn G. J. and Nelson W. (1970) A problem in the statistical comparison of measuring devices. *Technometrics*, **12**(1), 95–102.

Hansen M. H., Hurwitz W. N., and Bershad M. (1961) Measurement errors in censuses and surveys. *Bull. International Stat. Inst.*, **38**, 359–374.

Hanson R. H. and Marks E. S. (1958) Influence of the interviewer on the accuracy of survey results. *J. Amer. Stat. Assoc.*, **53**, 635–655.

Hudson W. R., Elkins G. E., Uddin W., and Reilley K. T. (1987) Improved Methods and Equipment to Conduct Pavement Distress Surveys, Final Report #FHWA-TS-87-213.

Humplick F. (1989) Theory and methods of analyzing infrastructure inspection output: Application to highway pavement surface condition evaluation. Ph.D. Dissertation, Civil Engineering Department, MIT, Cambridge, MA.

Jaech J. L. (1976) Large sample tests for Grubbs' estimators of instrument precision with more than two instruments. *Technometrics*, **18**(2), 127–133.

Jeyapalan K., Cable J. K., and Welper R. (1987) Automated pavement data collection equipment. Demonstration Project No. 72, Iowa DOT Evaluation of the PASCO Survey System.

Kish L. (1965) *Survey Sampling*. John Wiley & Sons, New York.

Maloney C. J. and Rastogi S. (1970) Significance test for Grubbs's estimators. *Biometrics*, **26**, 671–676.

Montgomery D. C. (1984) *Design and Analysis of Experiments*, Second Edition. John Wiley & Sons, New York.

Pindyck R. S. and Rubinfeld D. L. (1981) *Econometric Models & Economic Forecasts*, Second Edition. McGraw-Hill Book Company, New York.

Rummel R. J. (1970) *Applied Factor Analysis*. Northwestern University Press, Evanston, Illinois.

Shukla G. K. (1972)An invariant test for the homogeneity of variances in a two-way classification. *Biometrics*, **26**, 1063–1072.