Running Head: ML Estimation in VG

Maximum Likelihood Estimation in Validity Generalization with Examples

Nambury S. Raju

Illinois Institute of Technology

Fritz Drasgow

University of Illinois – Urbana/Champaign

David L. Blitz

Illinois Institute of Technology

and

Chicago School of Professional Psychology

Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Orlando, FL.

2003

Abstract

This paper briefly outlines the advantages of maximum likelihood (ML) estimation over other estimation methods. Then, the recently developed ML estimation procedures of Raju and Drasgow (2003) are described for use in validity generalization. Two examples are presented, comparing the traditional VG estimation methods based on the method of moments with the new ML estimation methods, across three different VG models/scenarios (bare-bones, use of artifact distributions, and direction corrections). The need for assessing the accuracy and comparability of the traditional and ML estimation procedures is addressed.

Maximum Likelihood Estimation in Validity Generalization with Examples

Since Schmidt and Hunter's seminal publication in 1977 (Schmidt & Hunter, 1977) on situational specificity, the assessment of the generalizability of organizational interventions has received a great deal of attention among researchers and practitioners, especially among industrial/organizational (I/O) psychologists concerning the generalizability of the validity of predictors across organizations. The art and science of validity generalization (VG) revolves around the estimation of the mean and variance of population validities. That is, given a set of k validity coefficients (correlations between the same or similar predictors and criteria in the same or similar jobs) obtained from samples drawn from k different populations, one is interested in obtaining an estimate of the mean and variance of population validities in order to establish whether the validity in question is (1) significant and substantial and (2) generalizable across populations. An estimate of the mean of population validities is used to answer the first question and an estimate of the variance of population validities is used to answer the second question of generalizability. The estimation of the mean and variance of population validities and the resulting substantive interpretations is the crux of VG research and practice.

There are currently several VG models and procedures for estimating the mean and variance of population validities. Some models/procedures are designed for use with observed correlations corrected for unreliability and range restriction, while others are not. Some models/procedures are couched in observed correlations, while others use transformations of such correlations to estimate the mean and variance of population validities. The statistical estimation methods employed also vary across procedures. For example, the early estimation procedures relied mostly on the method of moments, but some of the newer procedures use estimation methods based on the maximum likelihood principle. Recently, Bayesian estimation techniques have been used by VG researchers (Brannick, 2001). Statistical estimation procedures (e.g., the method of moments, maximum likelihood estimation, method of least squares, and Bayesian estimation) are described in standard texts on mathematical statistics (for example, Lehmann, 1983; Mood & Graybill, 1963; Rao, 1973).

The use of different statistical estimation methods has, at times, led to some confusion and controversy among VG researchers and practitioners. While the choice among the different statistical estimation methods may not be an easy one, especially for practitioners, the availability of procedures for estimating the VG parameters (mean and variance of population correlations) based on different statistical estimation methods would be very desirable. The choice of estimation method is usually based on a consideration of optimality. For example, ceteris paribus, an unbiased estimator should be preferred to a biased estimator. The early VG estimation procedures were based on the method of moments. The statistical properties of estimates based on the method of moments are not necessarily optimal. Maximum likelihood estimation is optimal in several important respects. For example, as shown by Kendall and Stuart (1977; 1979, pp. 38-81), maximum likelihood (ML) estimates are consistent and asymptotically efficient. Additional information about the ML estimates may be found in Lehmann (1983), Raju and Drasgow (2003), and Rao, (1973). These are briefly described in the paragraphs below.

Even though the ML estimation method has received substantial attention lately among VG researchers (Brannick, 2001; Erez, Bloom, & Wells, 1996; Hedges & Olkin, 1985), much of it is confined to Fisher's z-transformation of an observed correlation or validity coefficient. ML estimation procedures for VG parameters, when correlations are corrected for unreliability and range restriction, have not been previously derived. Given the many statistically optimal properties that ML estimators enjoy, it is natural and desirable to extend the ML estimation technique to VG situations in which correlations are corrected for unreliability in the predictor and criterion and direct range restriction on the predictor. Therefore, it is the purpose of this paper to briefly outline the recent work of Raju and Drasgow (2003) on the ML estimation in the context of VG. It should be noted that for the purposes of the flow of presentation and comprehensiveness, ML estimation methods are presented for both corrected and uncorrected correlations. Moreover, in this presentation we will only deal with correlation coefficients, not Fisher's z-transformations of such correlations. Readers interested in the use of Fisher's z transformation in VG analysis are referred to Erez et al. (1996) and Hedges and Olkin (1985).

Some Preliminaries

In this section, following Raju and Drasgow (2003), some well known results about correlations and a brief description of the fixed-effects and random-effects models will be presented. These models are receiving a great deal of attention lately, especially with respect to which VG parameters are estimated.

Single Population

Let the r_{xy} represent the correlation between a predictor (x) and a criterion (y) in a sample of size n drawn from a population. Let the correlation between x and y in the population be denoted as ρ_{xy}. Let the hypothesized relationship between r_{xy} and ρ_{xy} be expressed as:

$$r_{xy} = \rho_{xy} + e, \tag{1}$$

and let us further assume that r_{xy} is an unbiased estimate of ρ_{xy} and is normally distributed. Strictly speaking, r_{xy} is not an unbiased estimate of ρ_{xy}; in fact, it underestimates ρ_{xy} (Stuart & Kendall, 1977; Hedges & Olkin, 1985). According to Hedges (1989) and Hedges and Olkin (1985), this bias is seldom of practical significance if the sample size is not too small. Therefore, the current assumption may be restated as that the asymptotic (or large sample) distribution of r_{xy} is approximately normal with a mean of ρ_{xy} and an asymptotic variance of $\sigma^2_{r_{xy}}$. According to Kendall and Stuart (1977), the asymptotic mean/expectation and variance of r_{xy} can be expressed as:

$$E(r_{xy}) = \rho_{xy}, \tag{2}$$

$$\sigma^2_{r_{xy}} = \frac{(1-\rho^2_{xy})^2}{n}. \tag{3}$$

When r_{xy} is substituted for ρ_{xy} in Equation 3, this equation may be rewritten as:

$$\hat{\sigma}^2_{r_{xy}} = \frac{(1-r^2_{xy})^2}{n-1}. \tag{4}$$

<u>Several Populations</u>

Let us now consider *k* different populations drawn at random from a universe of populations and denote the observed correlation, based on a sample (of size n_i) drawn from population i, be denoted as $r_{x_i y_i}$ and the population correlation as $\rho_{x_i y_i}$. A major goal of VG methods is to estimate the mean and variance of population validities or correlations ($\rho_{x_i y_i}$). Another equally important goal is to develop appropriate confidence intervals for estimates of the mean. With this information at hand, researchers and practitioners can assess the degree to which validity is generalizable across organizations, provided that we continue to sample from the same universe of populations. Following Raju and Drasgow (2003), procedures for estimating the mean and variance of population validities are presented below using the well known ML approach. Prior to presenting these procedures, the fixed-effects and random-effects models are briefly described.

Fixed- and Random-Effects Models

In the meta-analysis/VG literature, a model is said to be fixed (or homogeneous) if $\rho_{x_i y_i}$ values are considered identical across the k populations (drawn at random from a universe of populations with a constant correlation) so that the variance of $\rho_{x_i y_i}$ values is zero. That is,

$$\rho_{x_1 y_1} = \rho_{x_2 y_2} = \ldots = \rho_{x_k y_k} = \rho . \tag{5}$$

In view of Equation 5, the fixed-effects model may be written as:

$$r_{x_i y_i} = \rho + e_i . \tag{6}$$

In Equations 5 and 6, ρ refers to the common population validity across the universe of populations. In the random-effects model, $\rho_{x_i y_i}$ values may not be equal across populations. That is,

$$r_{x_i y_i} = \mu_\rho + \tau_i + e_i , \tag{7}$$

where μ_ρ is the grand mean in the sense of the analysis of variance (ANOVA) terminology and $\tau_i = \rho_{x_i y_i} - \mu_\rho$. In the fixed-effects model, $\tau_i = 0$

for all i. In the random-effects model, $\sigma_\tau^2 = \sigma_\rho^2$, which is typically not equal to zero. Moreover, the magnitude of $\sigma_\tau^2 = \sigma_\rho^2$ is of critical importance in answering the question of whether validity is generalizable.

Maximum Likelihood Estimation with Uncorrected Correlations

In this section, we will describe procedures for estimating the relevant VG parameters when the observed correlations are not corrected for unreliability and/or range restriction. For ease of presentation, we will refer to this scenario as the 'bare-bones' VG model.

Fixed-Effect Model

In view of the fixed-effects VG model given in Equation 6, the (asymptotic) mean and variance of an observed correlation or validity coefficient may be expressed as:

$$E(r_{x_i y_i}) = \rho, \tag{8}$$

$$\sigma^2_{r_{x_i y_i}} = \sigma^2_{e_i}. \tag{9}$$

Using Equations 8-9, the ML estimate of ρ, under the assumption that the sample validities are normally distributed, may be written as:

$$\rho = \frac{\sum_{i=1}^{k} w_i r_{x_i y_i}}{\sum_{i=1}^{k} w_i}, \tag{10}$$

where

$$w_i = \frac{1}{\sigma^2_{e_i}}. \tag{11}$$

Equation 10 can be used to solve for ρ. According to Equation 10, the ML estimate of ρ is a weighted average of the observed correlations from k different populations. It should be noted that in a fixed-effects model, ρs do nor vary from population to population and, hence, there is no need to estimate their variance.

The weight for each validity coefficient is the reciprocal of its error variance. Typically, the bigger the sample size, the smaller the error variance and hence the bigger the weight. That is, validity coefficients based on bigger samples will receive bigger weights than validities based on smaller samples. In view of Equation 4, an estimate of each weight may be written as

$$\hat{w}_i = \frac{n_i - 1}{(1 - r_{x_i y_i}^2)^2} \qquad (12)$$

and used in Equation 10 to solve for ρ. It should be noted that Equations 10 and 11 are not new and were previously presented by Hedges and Olkin (1985); Erez et al. (1996) provided similar equations for Fisher's z transformation of an uncorrected correlation. According to Kendall and Stuart (1977), the maximum likelihood estimate $\hat{\rho}$ is (asymptotically) normally distributed with mean equal to ρ and sampling variance equal to

$$\sigma_{\hat{\rho}}^2 = \frac{1}{(\sum_{i=1}^{k} w_i)}. \qquad (13)$$

Using Equation 12, Equation 13 can be rewritten as:

$$\hat{\sigma}_{\hat{\rho}}^2 = \frac{1}{\sum_{i=1}^{k} \frac{(n_i - 1)}{(1 - r_{x_i y_i})^2}}. \qquad (14)$$

A confidence interval for ρ may be written as:

$$\hat{\rho} - a\hat{\sigma}_{\hat{\rho}} \prec \rho_{xy} \prec \hat{\rho} + a\hat{\sigma}_{\hat{\rho}}, \qquad (15)$$

where a is a constant and depends on the alpha level specified by the investigator. In practice, the population validity is assumed to be zero if this confidence interval contains the zero point. It should be noted that this error variance formula for $\hat{\rho}$ is only valid when the validity is the same across populations. Strictly speaking, one should first check to see if the hypothesis of equal population validities is valid, given the observed validities. Hedges and Olkin (1985) and Hunter and Schmidt (1990) offered statistical procedures (based on the chi-square statistic) for testing this null hypothesis.

Random-Effects Model

In view of Equation 7, the (asymptotic) mean and variance of an observed correlation may be written as:

$$E(r_{x_i y_i}) = \mu_\rho , \tag{16}$$

$$\sigma^2_{r_i y_i} = \sigma^2_\rho + \sigma^2_{e_i} . \tag{17}$$

A comparison of Equation 17 with Equation 9 shows an important distinction between the fixed-effects model and the random-effects model. The variance of population validities is included in the sampling variance of an observed correlation in the random-effects model, but not in the fixed-effects model. Therefore, the fixed-effects sampling variance of an observed correlation is smaller than or equal to the sampling variance in the random-effects model. This distinction is not always well understood by practitioners and has received a great deal of attention lately (Erez at al., 1996; Hedges & Vevea, 1998; Hunter & Schmidt, 2000; Overton, 1998).

Assuming that the sample validities are normally distributed, the ML estimates of the mean and variance of rhos (μ_ρ and σ^2_ρ) may be written as

$$\mu_\rho = \frac{\sum_{i=1}^{k} w_i r_{x_i y_i}}{\sum_{i=1}^{k} w_i}, \tag{18}$$

$$\sigma_\rho^2 = \frac{\sum_{i=1}^{k} w_i^2 [(r_{x_i y_i} - \mu_\rho)^2 - \sigma_{e_i}^2]}{\sum_{i=1}^{k} w_i^2}, \tag{19}$$

where

$$w_i = \frac{1}{\sigma_\rho^2 + \sigma_{e_i}^2}. \tag{20}$$

In view of Equation 4, an estimate of w_i ($\hat{w}_i$) can be written as:

$$\hat{w}_i = \frac{1}{\sigma_\rho^2 + \frac{(1 - r_{x_i y_i}^2)^2}{n-1}}. \tag{21}$$

These weights ($\hat{w}_i$) can then be used in solving Equations 18 and 19 simultaneously to obtain estimates for the mean and variance of population validities. It is not a difficult task, but numerical procedures are needed to successfully accomplish this estimation. Again, the sampling variance of $\hat{\mu}_\rho$ within the ML framework may be expressed as:

$$\hat{\sigma}_{\hat{\mu}_\rho}^2 = \frac{1}{(\sum_{i=1}^{k} \hat{w}_i)}. \tag{22}$$

Appropriate confidence intervals for μ_ρ can be developed using an equation similar to the one given in Equation 15.

Maximum Likelihood Estimation with Corrected Correlations

Up to this point, we have only looked at the relationship between observed validities and their population parameters. In validity generalization research, it is a common practice to correct observed validities for unreliability in the criterion and/or predictor reliability and range restriction on the predictor. Standard psychometric and statistical procedures (Lord & Novick, 1968) are used in making these corrections. Two well known procedures for making such corrections to observed validity coefficients are as follows: (1) Correcting correlations using hypothetical distributions of predictor reliability, criterion reliability and range restriction (or corrections based on hypothetical artifact distributions) and (2) correcting correlations at the study level (that is, using sample-based reliability and range restriction values). The currently available VG procedures for corrected correlations are based on one of these types of corrections. For ease of presentation, the psychometric model with corrections based on hypothetical artifact distributions will be hereafter referred to as VG with artifact distributions and the one based on study-level corrections as VG with direct corrections . Each of these models will be described below, along with procedures for obtaining the maximum likelihood estimates of the mean and variance of population validities (Raju & Drasgow, 2003). Since the random-effects model is considered by many to be a much more realistic model in practice, we will describe the ML estimation only for the random-effects model. Readers interested only in the fixed-effects model may ignore the estimation of the variance of population validities. For this group of researchers and practitioners, the estimation task will be rather straightforward and will not require any iterative algorithms.

VG with Artifact Distributions

Prior to defining this model, let $\rho_{x_i x_i}$ represent the unrestricted population reliability of x, $\rho_{y_i y_i}$ represent the unrestricted population reliability of y, and u_i represent the attenuated population range restriction factor or simply the ratio of attenuated, restricted population standard deviation to the attenuated, unrestricted population standard deviation. This model may be expressed as:

$$r_{x_i y_i} = \rho_{x_i y_i} \sqrt{\rho_{x_i x_i}} \sqrt{\rho_{y_i y_i}} \frac{u_i}{\sqrt{(1 + (u_i^2 - 1)\rho_{x_i y_i}^2 \rho_{x_i x_i} \rho_{y_i y_i}}} + e_i \qquad (23)$$

for study i. It should be noted that $\rho_{x_i y_i}$ is the unattenuated, unrestricted population validity (or correlation between x and y) in population i. For ease of presentation, the above equation may be rewritten as:

$$r_{x_i y_i} = h(\rho_{x_i y_i}, \rho_{x_i x_i}, \rho_{y_i y_i}, u_i) + e_i = h_i + e_i\,, \qquad (24)$$

where

$$h_i \equiv h(\rho_{x_i y_i}, \rho_{x_i x_i}, \rho_{y_i y_i}, u_i) = \rho_{x_i y_i} \sqrt{\rho_{x_i x_i}} \sqrt{\rho_{y_i y_i}} \frac{u_i}{\sqrt{(1 + (u_i^2 - 1)\rho_{x_i y_i}^2 \rho_{x_i x_i} \rho_{y_i y_i}}}\ . \qquad (25)$$

Using the ANOVA notation in Equation 7, Equation 24 may be rewritten as

$$r_{x_i y_i} = \mu_h + \tau_i + e_i\,, \qquad (26)$$

where μ_h is the grand mean and $\tau_i = h_i - \mu_h$. In the fixed-effects model, $\tau_i = 0$ for all i. In the random-effects model, $\sigma_\tau^2 = \sigma_h^2$. In view of Equation 24, at an individual study level, the mean (or expectation) and variance of $r_{x_i y_i}$ may be written, respectively, as:

$$E(r_{x_i y_i}) = \mu_h\,, \qquad (27)$$

$$\sigma^2_{r_i y_i} = \sigma^2_h + \sigma^2_{e_i} . \tag{28}$$

As before, the maximum likelihood estimates of μ_h and σ^2_h can be obtained by solving the following two equations simultaneously.

$$\mu_h = \frac{\sum_{i=1}^{k} w_i r_{x_i y_i}}{\sum_{i=1}^{k} w_i} , \tag{29}$$

$$\sigma^2_h = \frac{\sum_{i=1}^{k} w_i^2 [(r_{x_i y_i} - \mu_h)^2 - \sigma^2_{e_i}]}{\sum_{i=1}^{2} w_i^2} . \tag{30}$$

As in the previous scenarios and in view of Equation 4, an estimate of w_i ($\hat{w}_i$) can be written as:

$$\hat{w}_i = \frac{1}{\sigma^2_h + \frac{(1 - r^2_{x_i y_i})^2}{n-1}} . \tag{31}$$

Equations 29-30 form the basis for iteratively solving for the maximum likelihood estimates of the mean and variance of h. Finally, the sampling variance of $\hat{\mu}_h$ within the ML framework may be expressed as:

$$\hat{\sigma}^2_{\hat{\mu}_h} = \frac{1}{(\sum_{i=1}^{k} \hat{w}_i)} . \tag{32}$$

While solving Equations 29-30 will yield maximum likelihood estimates of μ_h and σ^2_h, these are not the parameters of interest in a VG analysis because they refer to the restricted, attenuated relationship. The parameters of interest are μ_ρ and σ^2_ρ. Due to the invariance property of ML estimators, it is possible to obtain the ML estimators for

μ_ρ and σ_ρ^2 with the following equations (Raju & Drasgow, 2003).

$$\hat{\mu}_\rho = \frac{\hat{\mu}_h}{\sqrt{\mu_{\rho_{xx}}\mu_{\rho_{yy}}(\mu_u^2 + \hat{\mu}_h^2(1-\mu_u^2))}}, \tag{33}$$

$$\hat{\sigma}_\rho^2 = \frac{\hat{\sigma}_h^2 - (B^2\sigma_{\rho_{xx}}^2 + C^2\sigma_{\rho_{yy}}^2 + D^2\sigma_u^2)}{A^2}. \tag{34}$$

The terms A, B, C, and D are the partial derivatives of h (Equation 25) with respective to $\rho_{x_iy_i}$, $\rho_{x_ix_i}$, $\rho_{y_iy_i}$, and u_i, respectively. The sampling variance of $\hat{\mu}_h$ can be obtained with

$$\hat{\sigma}_{\hat{\mu}_\rho}^2 = \frac{\left(\mu_u^2\mu_{\rho_{xx}}\mu_{\rho_{yy}}\right)^2}{\left(\mu_{\rho_{xx}}\mu_{\rho_{yy}}(\mu_u^2 + \hat{\mu}_h^2(1-\mu_u^2)\right)^3}\frac{1}{(\sum_{i=1}^{k}\hat{w}_i)}, \tag{35}$$

where $\hat{w}_i$ are as defined in Equation 31. Please refer to Raju and Drasgow (2003) for additional details.

The statistical rationale used here for estimating the mean and variance of ρ is similar to the one used by Raju and Burke (1983) in developing their TSA procedures. It may be possible to use any one of the six currently available procedures (Callender & Osborn; 1980, Pearlman et al., 1980; Raju & Burke, 1983; Schmidt et al., 1980; Law et al., 1994) for this purpose. Among the six procedures, the TSA1 and TSA 2 procedures of Raju and Burke (1983) make direct use of the mean and variance of hypothetical range restriction values while the other four procedures use the mean and variance of a function of range restriction values. Given this relative simplicity associated with the TSA procedures, we have decided to use some of the same techniques used in one of the TSA procedures (TSA 1) to estimate the mean and variance of ρ from the maximum likelihood estimates of the mean and variance of h. As the acronym implies, this procedure uses the Taylor series approximation (TSA) to estimate the mean and variance of a function of several variables. These are asymptotic estimates (Kendall & Stuart, 1977); that is, as the sample size increases, these estimates converge to their population parameters.

The reliabilities and range restriction values appearing in Equations 23-25 are commonly referred to as artifacts. Their means and variances in Equations 33-35 are assumed known at the population level. Furthermore, these reliabilities are for the unattenuated and unrestricted populations, whereas the range restriction values are for the attenuated (adjusted for unreliability) populations. The prevailing practice (Callender & Osborn; 1980, Pearlman et al., 1980; Raju & Burke, 1983; Schmidt et al., 1980; Law et al. 1994) is to use hypothetical distributions (means and variances) of artifacts in estimating the mean and variance of unattenuated and unrestricted population validities.

Within the framework of VG with artifact distributions, the use of hypothetical artifact distributions gets around the frequently faced problem of not having access to sample-based reliability and range restriction values. In practice, one should be careful, however, in accepting the validity of hypothetical artifact distributions across all conditions. Hypothetical distributions of artifacts are probably justifiable in some situations, but may not be in others. Practitioners need to pay careful attention to the degree to which hypothetical distributions match the true distributions of artifacts.

In addition to the assumption about the validity of artifact distributions, the currently available procedures for VG with artifact distributions assume that the hypothetical distributions are uncorrelated across populations. Several investigators have commented on the meaning and tenability of this assumption. For this assumption to be true, according to James, Demaree, Mulaik, and Ladd (1992), the situational moderators that underlie situational specificity must be independent from the statistical artifacts. If the two are not independent, then the current procedures remove part of the actual variance attributable to the situational moderators when they subtract the "artifactual" variance. In two recent Monte Carlo studies, Raju et al. (1998) and Thomas and Raju (1998) have shown that if this assumption is violated, then the σ_{ρ}^{2} estimates derived from the procedures based on bare-bones VG are not always accurate. In response to the problems associated with the use of hypothetical artifact distributions and the uncorrelatedness of these distributions, Raju, Burke, Normand, and Langlois (1991) proposed a different model (VG with direct corrections), which assumes that the study-level range restriction and reliability values are available. Then they go on to show that this model may still be useful when only partial (study-level) artifact data are available. Following Raju and Drasgow (2003), a description of this model is presented below along with its maximum likelihood estimates of the mean and variance of unrestricted and unattenuated population validities.

VG with Direct Corrections

As before, let $r_{x_i y_i}$ represent the observed correlation between x and y in a sample drawn from population i. Let $r_{x_i x_i}$ and $r_{y_i y_i}$ represent the sample-based (restricted) reliabilities for x and y, respectively. Finally, let u_i represent the unattenuated, but sample-based range restriction value. Then, using classical test theory (Lord & Novick, 1968), a sample-based estimate of the unrestricted and unattenuated population validity (ρ_i) may be written as:

$$r_i = \frac{g_i r_{x_i y_i}}{\sqrt{r_{x_i x_i} r_{y_i y_i} - r^2_{x_i y_i} + g^2_i r^2_{x_i y_i}}}, \tag{36}$$

where $g_i = 1/u_i$. In view of this equation, the model for VG with direct corrections (within the random-effects framework) may be expressed as:

$$r_i = \mu_\rho + \tau_i + e_i . \tag{37}$$

As previously defined, μ_ρ is the grand mean and $\tau_i = \rho_i - \mu_\rho$. In the fixed-effects model, $\tau_i = 0$ for all i; in the random-effects model, $\sigma^2_\tau = \sigma^2_\rho$. In view of Equation 37, at an individual study level, the mean (or expectation) and variance of r_i may be written, respectively, as:

$$E(r_i) = \rho_i , \tag{38}$$

$$\sigma^2_{r_i} = \sigma^2_\rho + \sigma^2_{e_i} . \tag{39}$$

According to Raju et al. (1991), an asymptotic variance of e_i may be written as:

$$\hat{\sigma}^2_{e_i} = \frac{g^2_i r_{x_i x_i} r_{y_i y_i} (r_{x_i x_i} - r^2_{x_i y_i})(r_{y_i y_i} - r^2_{x_i y_i})}{(n_i - 1)\hat{V}^3_i}, \tag{40}$$

where

$$\hat{V}_i = r_{x_i x_i} r_{y_i y_i} - r^2_{x_i y_i} + g^2_i r^2_{x_i y_i} . \tag{41}$$

Given this model, the necessary maximum likelihood equations for VG with direction corrections may be expressed as:

$$\mu_\rho = \frac{\sum_{i=1}^{k} w_i r_i}{\sum_{i=1}^{k} w_i}, \tag{42}$$

$$\sigma_{\rho}^{2} = \frac{\sum_{i=1}^{k} w_i^2 [(r_i - \mu_{\rho})^2 - \sigma_{e_i}^2]}{\sum_{i=1}^{k} w_i^2}, \tag{43}$$

where an estimate of w_i ($\hat{w}_i$) can be written as:

$$\hat{w}_i = \frac{1}{\sigma_{\rho}^{2} + \frac{g_i^2 r_{x_i x_i} r_{y_i y_i} (r_{x_i x_i} - r_{x_i y_i}^2)(r_{y_i y_i} - r_{x_i y_i}^2)}{(n_i - 1)\hat{V}_i^3}}. \tag{44}$$

Equations 42-43 will need to be solved iteratively for obtaining the maximum likelihood estimates of the mean and variance of ρ. As in the previous models, the sampling variance of $\hat{\mu}_{\rho}$ within the ML framework may be expressed as:

$$\hat{\sigma}_{\hat{\mu}_{\rho}}^{2} = \frac{1}{(\sum_{i=1}^{k} \hat{w}_i)}. \tag{45}$$

A Few Examples

We have briefly outlined the advantages of ML estimation methods over other estimation methods (for example, the method of moments) and then presented Raju and Drasgow's (2003) procedures for estimating the mean and variance of population validities within this framework. ML-based VG estimates are presented for three different scenarios: bare-bones (uncorrected correlations), artifact distributions (correlations corrected with hypothetical artifact distributions), and direct corrections (correlations corrected with sample-based artifact data at the study level). All three models are expressed in terms of observed correlations without any transformations.

Although it is aesthetically pleasing to derive new estimators for important VG parameters, it is also important to consider the "So what?" question. Specifically, do the

ML estimators described herein for VG with artifact distributions and VG with direct corrections provide improved estimation of the mean and variance of population correlations? One clear benefit is that we know the sampling distribution of ML estimates (i.e., they are asymptotically normal); consequently we can legitimately construct confidence intervals for VG parameters based on our sample estimates.

There is a pressing need for assessing empirically the accuracy of the ML estimates for the three models studied here. In addition to accuracy, there is a need for assessing the relative accuracy of the ML estimates and the estimates from other methods, such as the method of moments. Large scale Monte Carlo studies are needed for a comprehensive assessment of the accuracy and utility of these estimations. Such studies would reveal whether the ML estimates are more precise for the types of data typically analyzed by applied psychologists.

We are currently in the process of starting such an investigation. We have recently completed the development of a computer program in FORTRAN for determining the necessary ML estimates for the three models. The iterative computations involved in the ML estimation are carried out with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Press, Flannery, Teukolsky, & Vetterling, 1986). The BFGS algorithm improves upon the well-known Davidon-Fletcher-Powell algorithm for solving non-linear simultaneous equations. We will use these programs to illustrate what the ML estimates are like and how they compare with the traditional (currently used) VG estimates. Two simple examples are used for this purpose. It should be noted that there is really nothing unique about these examples except that they are used here for illustrative purposes.

Both VG examples consist of 10 validity studies. In Example 1 (Table 1), there are no sample-based artifact data, whereas, in Example 2 (Table 2), partial, sample-based range restriction, predictor reliability, and criterion reliability data are assumed known. In both examples, VG estimates for the three models are reported separately for the traditional and ML estimation methods. Each set of estimates consists of the mean of rhos ($\hat{\mu}_\rho$), variance of rhos ($\hat{\sigma}^2_\rho$), and the fixed and random sampling variances ($\hat{\sigma}^2_{\hat{\mu}_\rho}$) of estimate of the mean of rhos. The top half of each table shows the actual correlations used, and the bottom half shows the traditional and ML estimates.

<u>Example 1</u>

In the first example (Table 1), the bare-bones mean estimates are .330 and .335 for the traditional and ML estimates, respectively. The traditional mean estimates for the other two models (.730 and .649, respectively) are substantially higher. This is an expected result because the observed correlations are corrected for range restriction and unreliability in these models. In the model with artifact distributions, the artifact distributions previously recommended by Pearlman et al. (1980) were used. The mean values from the same artifact distributions were also used in the model with direct corrections; this is consistent with Raju et al.'s (1991) recommendation when sample-based artifact information is completely unavailable. Compared to the traditional estimates, the ML estimates are slightly, but consistently higher; in the model with direct corrections, the ML estimate appears to be substantially higher than the traditional estimate (.735 vs. .649).

According to the data in Table 1, the variance estimates appear to be different across the three models and also across the two estimation methods. The traditional variance estimates for the three models are .016, .037, and .027, respectively. Similarly, the ML variance estimates for the three models are .018, .042, and .018. The difference between the two estimation methods, both in terms of means and variances, is the largest for the VG model with direct corrections. The fixed and random sampling variances appear to be comparable across both methods of estimation, with the VG model with artifact distributions showing the biggest change (from .005 to .001). The ML estimates of the sampling variance appear to be about the same or slightly smaller than the traditional estimates.

Example 2

Data from the second example are shown in Table 2. The top half of this table shows not only the observed correlations but also the sample-based range restriction and predictor and criterion reliability values. It should be noted that only partial information is assumed known for the latter three indices. It should also be noted that these partial data were used in the VG model with direct corrections. In general, as shown at the bottom of Table 2, the ML estimates of the mean and variance of rhos appear to be slightly larger than the traditional estimates. With respect to the fixed and random sampling variances, both methods of estimation appear to give similar results across the three models. Finally, the mean estimates are higher for the VG model with artifact distributions than for the VG model with direct corrections; the opposite appears to be true for the variance estimates. Some or all of this variability may be due to the fact that the means and variances of the sample-based artifact distributions are different from the means and variances of Pearlman et al.'s distributions of artifacts.

In summary, the data in Tables 1 and 2 are just two illustrations of what the estimates from the traditional and ML methods may look like. The traditional and ML estimates appear to be different at times, which may be unique for these two data sets used. These are some of the examples we used in debugging the ML estimation program. After some more checks with the program, we will be ready for a Monte Carlo study to asses the comparability and accuracy of the traditional and ML estimation methods. Other than the known, desirable statistical properties (asymptotic consistency and efficiency) associated with ML estimations, nothing can be said at this time about the degree of the accuracy of the ML estimates. We need a comprehensive Monte Carlo investigation to answer that question.

References

Brannick, M. T. (2001). Implications of empirical Bayes meta-analysis for test validation. Journal of Applied Psychology, 86, 468-480.

Callender, J.C. & Osburn, H. (1980). Development and test of a new model for validity generalization. Journal of Applied Psychology, 65, 543-558.

Erez, A., Bloom, M.C., & Wells, M.T. (1996). Using random than fixed effects models in meta-analysis: Implications for situational specificity and validity generalization. Personnel Psychology, 49, 275-306.

Hedges, L.V. (1989). An unbiased correlation for sampling error in validity generalization studies. Journal of Applied Psychology, 74, 469-477.

Hedges, L.V., & Olkin, I. (1985). Statistical methods for meta-analysis. New York: Academic Press.

Hedges, L.V., & Vevea, J.L. (1998). Fixed and random effects models in meta-analysis. Psychological Methods, 3, 486-504.

Hunter, J.E., & Schmidt, F.L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. International Journal of Selection and Assessment, 8, 275-292.

Hunter, J.E., & Schmidt, F.L. (1990). Methods of meta-analysis. Beverley Hills, CA: Sage Publications.

James, L.R. J., Demaree, R.G., Mulaik, S.A., & Ladd, R.T. (1992). Validity generalization in the context of situational models. Journal of Applied Psychology, 77, 3-14.

Kendall, M., & Stuart, A. (1973). The advanced theory of statistics (Volume 2, 3rd Edition). New York: Macmillan.

Kendal, M., & Stuart, A. (1977). The advanced theory of statistics (Volume 1, 4th Edition). New York: Macmillan.

Law, K., Schmidt, F.L., & Hunter, J.E. (1994). Nonlinearity of range restriction in meta-analysis: Test of an improved procedure. Journal of Applied Psychology, 79, 425-438.

Lehmann, E.L. (1983). Theory of point estimation. New York: Wiley.

Lord, F.M., & Novick, (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Mood, A.M., & Graybill, F.A. (1963). Introduction to the theory of statistics (2nd Edition). NY: McGraw-Hill.

Overton, R.C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. Psychological Methods, 3, 354-379.

Pearlman, K., Schmidt, F.L. & Hunter, J.E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. Journal of Applied Psychology, 65, 373-406.

Press, W.H., Flannery, B.P., Teukolsky, S.A., & Vetteling, W.T. (1986). Numerical recipes: The art of scientific computing. Cambridge, England: Cambridge University Press.

Raju, N.S., Anselmi, T.V., Goodman, J.S., & Thomas, A. (1998). The effect of correlated artifacts and true validity on the accuracy of parameter estimation in validity generalization. Personnel Psychology, 51, 453-465.

Raju, N.S. & Burke, M.J. (1983). Two new procedures for studying validity generalization. Journal of Applied Psychology, 68, 382-395.

Raju, N.S., Burke, M.J., Normand, J., & Langlois, G.M. (1991). A new meta-analytic approach. Journal of Applied Psychology, 76, 432-446.

Raju, N.S., & Drasgow, F. (2003). Maximum likelihood estimation in validity generalization. In K.R. Murphy (ed.), Validity generalization: A critical review, (pp. 263-285). Mahwah, NJ: Erlbaum.

Rao, C.R. (1973). Linear statistical inference and its application (2nd Edition). NY: John Wiley.

Schmidt, F.L., Hunter, J.E. (1977). Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 62, 529-540.

Schmidt, F.L., Hunter, J.E., & Raju, N.S. (1988). Validity generalization and situational specificity: A second look at the 75% rule and the Fisher's z transformation. Journal of Applied Psychology, 73, 665-672.

Schmidt, F.L., Gast-Rosenberg, I., & Hunter, J.E. (1980). Test of a model of validity generalization: Results for computer programmers. Journal of Applied Psychology, 65, 643-661.

Thomas, A.L. & Raju, N.S. (1998, April). An examination and evaluation of James et al. (1992) VG estimation procedure when artifacts are corrected. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.

Table 1

Example 1: A data set of 10 correlations with no sample-based artifact data

Study	Sample Size	Correlation
1	75	.140
2	75	.250
3	75	.400
4	100	.300
5	100	.510
6	100	.270
7	100	.620
8	80	.300
9	80	.350
10	80	.050
Mean	86.5	.330
Variance		.025

Results from the three models

Model	Traditional Estimates				ML Estimates			
	Mean ($\hat{\mu}_\rho$)	Variance ($\hat{\sigma}^2_\rho$)	Fixed ($\hat{\sigma}^2_{\hat{\mu}_\rho}$)	Random ($\hat{\sigma}^2_{\hat{\mu}_\rho}$)	Mean ($\hat{\mu}_\rho$)	Variance ($\hat{\sigma}^2_\rho$)	Fixed ($\hat{\sigma}^2_{\hat{\mu}_\rho}$)	Random ($\hat{\sigma}^2_{\hat{\mu}_\rho}$)
Bare-Bones	.330	.016	.001	.003	.335	.018	.001	.003
With Artifact Distributions	.730	.037	.001	.005	.739	.042	.000	.001
With Direct Corrections	.649	.027	.003	.005	.735	.018	.001	.004

Table 2

Example 2: A data set of 10 correlations with partial sample-based artifact data

Study	Sample Size (N)	Correlation (r)	Range Restriction (u)	Predictor Reliability (r_{xx})	Criterion Reliability (r_{yy})
1	60	.140		.850	.600
2	75	.250	.800		.700
3	85	.450	.880	.830	
4	110	.320			.820
5	50	.410		.790	
6	90	.600	.900		
7	100	.620			
8	65	.350	1.000	.900	.590
9	80	.350	.500	.850	.650
10	65	.190	.900	.760	.640
Mean	78.0	.338	.825	.832	.689
Variance		.016	.025	.002	.006

<u>Results from the three models</u>

Model	Traditional Estimates				ML Estimates			
	Mean ($\hat{\mu}_\rho$)	Variance ($\hat{\sigma}^2_\rho$)	Fixed ($\hat{\sigma}^2_{\hat{\mu}_\rho}$)	Random ($\hat{\sigma}^2_{\hat{\mu}_\rho}$)	Mean ($\hat{\mu}_\rho$)	Variance ($\hat{\sigma}^2_\rho$)	Fixed ($\hat{\sigma}^2_{\hat{\mu}_\rho}$)	Random ($\hat{\sigma}^2_{\hat{\mu}_\rho}$)
Bare-Bones	.338	.006	.001	.002	.350	.010	.001	.002
With Artifact Distributions	.746	.002	.001	.001	.767	.011	.000	.001
With Direct Corrections	.526	.014	.002	.003	.561	.017	.002	.004