Maximum Likelihood Estimation

Overidentified models have more solvable equations than unknown parameters, but the best estimate of the value can be obtained through an iterative process. In practice, correlations are not typically used to derive factor loadings or path coefficients. The primary reason we do not typically use correlations to derive path estimates in practice is that correlations use standardized variables and important information about the variances of the variables is lost (i.e., the variances of each of the variables is assumed to be 1). It can also be shown that for many different factor models, use of correlations can lead to erroneous results (Cudek, 1989). Thus, covariances, the unstandardized version of correlations, are used.

The logic of deriving estimates of the loadings remains the same. There are a set of equations that describe the model (i.e., *structural equations*) and some known values about the relations among all of the variables (i.e., a matrix of covariances). For complicated models, the most common approach to solving the set of structural equations is through a calculus-based method called *maximum likelihood* (ML). Maximum likelihood solves for the loadings by minimizing the discrepancy between the equations implied by the model and the obtained covariances.[1] This discrepancy is mathematically described as:

$$S - \hat{\Sigma}(\theta)$$

Where *S* is the covariance matrix obtained from the data, and $\hat{\Sigma}(\theta)$ is matrix notation for a covariance matrix implied by the hypothesized model.

Certain values for the relations among the variables are implied by certain specified models. We can examine the fit of the hypothesized model to the data, by comparing the implied (or "reproduced") covariances to those obtained. The ML solution is obtained by minimizing the following (somewhat frightening) *fit function*:

$$F_{ML} = \log\left|\hat{\Sigma}(\theta)\right| + tr\left(S\hat{\Sigma}^{-1}(\theta)\right) - \log|S| - (p+q)$$

log is the natural logarithm function (base *e*), $\hat{\Sigma}(\theta)$ is the covariance matrix implied by the model, *S* is the observed covariance matrix, *tr* is the trace matrix algebra function, and *(p + q)* is equal to the number of coefficients that need to be estimated in the model. The superscript in the middle, $^{-1}$, is the inverse matrix function, so the inverse of the implied matrix must be possible.[2]

The maximum likelihood estimator is *asymptotically unbiased*, meaning that in larger samples it is an unbiased estimation of the population value. It is an efficient estimator, so it provides a variance estimate that is smaller than other consistent estimation

---

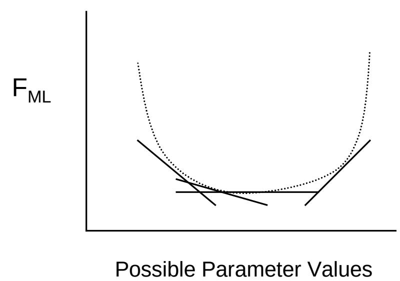
[1] The word "maximum" is used because the process maximizes the joint probability density function for the function or the parameters being estimated.

[2] Computer packages sometimes print an error message stating that the "inverse of sigma is not positive definite." The message is in reference to the assumption that the variance-covariance matrix must be positive definite—the inversion must be possible. This indicates a severe problem with the model because one or more of the implied variances from the variance/covariance matrix is negative.

methods. ML is asymptotically normal, so allows for convenient statistical tests like the Wald ratio test. With simple structural equation models, such as a multiple regression, and normal distribution assumptions, the maximum likelihood estimate will provide identical values to the OLS regression. ML, however, is a more general method that can be used with much more complicated models.

Fit, Chi-square, and df

ML is an iterative process, so initial starting values (i.e., guesses) are generated by the computer, the discrepancy between the implied and the obtained covariance matrix is computed, then new guesses are entered, and so on, until the minimum possible discrepancy values is obtained. Each step is called an *iteration*. The idea is similar to the idea of ordinary least squares (OLS) in regression in which the squared errors or residuals are minimized to obtain the best fit of the regression line to the data and the regression coefficients.



To find the minimum value of the F_{ML} discrepancy (fit) function, derivatives (i.e., second partial derivatives with respect to variance-covariance values) are used to draw tangent lines that correspond a point on the curve. When the tangent line has a slope of zero, the minimized value of the function has been found. The computer stops and generates values for the fit of the overall model and the parameter values. The final value can be used in a chi-square test [$\chi^2 = (N-1)F_{ML}$]. If the fit is perfect, there will be no discrepancy between the implied and obtained covariances, and the chi-square will be zero. A chi-square nonsignificantly different from zero indicates a good fit. Significantly positive chi-squares indicate poor fit.

Positive degrees of freedom do not guarantee all aspects of the model are identified and a solution can be found, but *df* > 0 is required for identification of the model. It should be noted that most textbooks give the following formula (or a variation using *p* + *q* to distinguish paths between exogenous and endogenous variables from those between endogenous paths).

$$df = \frac{v(v+1)}{2} - p$$

This formula is used because the number of unique variance/covariance elements (including the diagonal) is *v(v+1)*. Using this method, however, means that one must count the number of variances in the model when determining the value of *p*. Both models lead to the same result.

References and Further Reading

Bollen, K.A. (1989). Structural Equations with Latent Variables. New York: John Wiley. ISBN: 0-4710-1171-1.
Cudek, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 2,* 317-327.
Eliason, S.R. (1993). *Maximum likelihood estimation: Logic and practice.* Newbury Park, CA: Sage. QASS #96.
Mulaik, S.A. (2009). *Linear causal modeling with structural relations.* Boca Raton, FL: Chapman & Hall/CRC.