ON THE ASYMPTOTIC BIAS OF ESTIMATORS UNDER PARAMETER DRIFT

A. SHAPIRO and M.W. BROWNE

University of South Africa, P O Box 392, Pretoria 0001, South Africa

Abstract It is shown that, under a natural assumption, minimum discrepancy estimators in the analysis of moment structures are asymptotically unbiased

Keywords: asymptotic bias, asymptotic distribution, moment structures, parameter drift

1. Introduction

It is common practice to ensure the existence of an asymptotic distribution of a test statistic under the alternative hypothesis by introducing an assumption of parameter drift (Wald, 1943; Stroud, 1972). This assumption is not essential for the existence of an asymptotic distribution of estimators, but does result in a simplification of the expression for the asymptotic covariance matrix of estimators.

Recently Bentler and Dijkstra (1985) investigated the asymptotic distributions of estimators and test statistics under structural equation models. Assuming parameter drift, they showed that generalized least squares estimators are asymptotically normal and gave a corresponding "asymptotic bias" (Bentler and Dijkstra, 1985, equation (1.4.6); Bentler, 1985, p. 45). At first sight this appears to conflict with the known result (e.g. Dijkstra, 1983, Theorem 3) that, under the alternative hypothesis, these estimators are asymptotically unbiased for their population counterparts without parameter drift.

The expression for bias under population drift (Bentler and Dijkstra, 1985, equation (1.4.6)) is indeterminate. It involves an arbitrarily chosen vector, δ, which specifies the drift process and cannot be estimated in practice. Our aim is to point out that if the population value of the parameter vector is chosen appropriately, then constraints are imposed on the drift vector, δ, and the estimator is *asymptotically unbiased*. It is argued further that this choice of the population value to be estimated is a natural one.

2. Structural models and estimation

Let s be an $m \times 1$ vector statistic based on a sample of size n. It will be assumed that s is a consistent estimator of the population value σ^* of a parameter vector σ and that $n^{1/2}(s - \sigma^*)$ is asymptotically normal with a null mean vector and an $m \times m$ covariance matrix Γ. For example, in the analysis of covariance structures, s and σ usually represent elements of the sample covariance matrix S and population covariance matrix Σ, respectively. Validity of the assumptions is then ensured by the Law of Large Numbers and the Central Limit Theorem.

A structural model for the parameter vector σ is given by a vector valued function $\sigma = \sigma(\theta)$ of a $q \times 1$ vector θ. For the sake of simplicity we suppose that the model is unconstrained, i.e. there are no restrictions on the parameter vector θ. The model is said to hold if there exists a θ_0 such that $\sigma^* = \sigma(\theta_0)$. In this case we also say that the model is correct. An estimator $\hat{\theta}$ of θ_0 is obtained as a minimizer of the function $F(s, \sigma(\cdot))$. Here $F(s, \sigma)$

is a nonnegative, twice continuously differentiable function of two vector variables, s and σ, such that $F(s, \sigma)$ is zero if and only if $s = \sigma$. In the terminology of Browne (1982), F is called a discrepancy function. Two discrepancy functions which are often employed in computer programs for the analysis of covariance structures are

$$F_{ML}(S, \Sigma) = \log|\Sigma| - \log|S| + \mathrm{tr}(S\Sigma^{-1}) - p,$$

which corresponds to the method of maximum likelihood under normality assumptions, and

$$F_{GLS}(S, \Sigma) = \tfrac{1}{2}\mathrm{tr}\left[(S - \Sigma)S^{-1}\right]^2.$$

This discrepancy function F_{GLS} is an example of a generalized least squares discrepancy function and may be expressed in the form

$$F(s, \sigma \mid W) = (s - \sigma)'W(s - \sigma),$$

which is employed in Bentler and Dijkstra (1985).

The estimator $\hat{\theta}$ is a function of the statistic s and can be calculated even when the model is not correct. It is then of interest to investigate statistical properties of $\hat{\theta}$. Let us consider a minimizer θ_0 of the function $F(\sigma^*, \sigma(\cdot))$ associated with the population value σ^* of σ. Suppose that θ_0 exists, is *unique*, and is *independent* of the sample size n. Then under mild additional conditions, it can be shown that $\hat{\theta}$ is a *consistent* and asymptotically *unbiased* estimator of θ_0. Specifically suppose that all functions involved in the model $\sigma = \sigma(\theta)$ are twice continuously differentiable, and let H_{ss}, $H_{s\theta}$ and $H_{\theta\theta}$ be the Hessian matrices of second-order partial derivatives of $F\{s, \sigma(\theta)\}$ calculated at the point (σ^*, θ_0), e.g. $H_{s\theta} = (\partial^2/\partial s \partial \theta')F(\sigma^*, \sigma(\theta_0))$. We assume that the matrix $H_{\theta\theta}$ is non-singular. Then the following result is a consequence of the Implicit Function Theorem (see Shapiro, 1983, Theorem 5.4; Dijkstra, 1983, Theorem 3).

Proposition 1. *When σ^* is independent of n, the asymptotic distribution of $n^{1/2}(\hat{\theta} - \theta_0)$ is multivariate normal with a null mean vector and covariance matrix*

$$\Pi = H_{\theta\theta}^{-1}H_{s\theta}'\Gamma H_{s\theta}H_{\theta\theta}^{-1}. \qquad \square \qquad (1)$$

Similar results on asymptotic normality of estimators have been derived by various authors in the context of misspecified models (see Huber (1967), Foutz and Srivastava (1977), White (1982), and references therein). There θ_0 is the parameter vector which maximizes the expected value of a quasi-log-likelihood function. For example, in the analysis of covariance structures the normal theory maximum likelihood approach leads to the minimizer θ_0 of the function

$$\log|\Sigma(\theta)| + \mathrm{tr}\left[\Sigma^*\{\Sigma(\theta)\}^{-1}\right]$$

where Σ^* is the expected value of S. This agrees with our definition of θ_0 corresponding to the discrepancy function F_{ML}.

Proposition 1 shows that $\hat{\theta}$ is an asymptotically *unbiased* estimator of the minimizer θ_0 of $F(\sigma^*, \sigma(\cdot))$. When the model is not correct and $\sigma^* \neq \sigma(\theta_0)$ the "population" value of θ is not unambiguously defined. However, since $\hat{\theta}$ is a consistent and asymptotically unbiased estimator of θ_0, the latter would seem to be the natural candidate for the "population" value of θ to be estimated. Notice that if the model is not correct, the value of θ_0 will depend on the choice of the discrepancy function F and can vary from one discrepancy function to another.

Up to this point, the population value σ^* of σ has been assumed to be constant (independent of sample size, n). Under this assumption it was possible to give (Proposition 1) the asymptotic distribution of the estimator $\hat{\theta}$ when the model is not correct. The minimum discrepancy test statistic $nF(s, \sigma(\hat{\theta}))$ will, however, not have an asymptotic distribution under these circumstances. An asymptotic distribution for the test statistic is desirable to be used as an approximation in practical applications. In order to obtain this, another assumption limiting the size of the systematic error $\sigma^* - \sigma(\theta_0)$, relative to the sampling error in s, is necessary. A commonly employed assumption (cf. Stroud, 1972; Shapiro, 1983; Browne, 1984) is the assumption of "population drift" first used in Wald (1943). We now regard the population value $\sigma^* = \sigma_n$ as being a *function of sample size* which converges at a rate of $O(n^{-1/2})$ to a point $\sigma_0 = \sigma(\theta_0)$ satisfying the model. An assumption of this type is made in Bentler and Dijkstra (1985, p. 14).

Essentially they assume that

$$\sigma_n = \sigma_0 + n^{-1/2}\delta + o(n^{-1/2}), \qquad (2)$$

where δ is a fixed vector.

It was mentioned earlier that the assumption of population drift is not necessary to ensure the existence of an asymptotic distribution of the estimator $\hat{\theta}$. For ensuring an asymptotic distribution of the test statistic, the particular choice of δ in (2) is not important. An arbitrary choice of δ will, however, imply a redefinition of the parameter to be estimated and result in the asymptotic bias given in Bentler and Dijkstra (1985). This conflicts with the conclusion of unbiasedness implied by Proposition 1 where no population drift is assumed.

It seems natural to define population drift in such a manner that the parameter θ_0 to be estimated does not change when sample size is changed. Let $\theta_{0,n}$ be the minimizer of $F(\sigma_n, \sigma(\cdot))$. We let the drift occur in such a manner that

$$\theta_{0,n} = \theta_0 \qquad (3)$$

for all n. Thus the minimizer $\theta_{0,n}$ is fixed and independent of n. It follows from Taylor's theorem and the second order structure of discrepancy functions (Shapiro, 1985) that

$$\theta_{0,n} = \theta_0 + (\Delta'V\Delta)^{-1}\Delta'V(\sigma_n - \sigma_0) + \varepsilon_n, \qquad (4)$$

where $\Delta = (\partial/\partial\theta')\sigma(\theta_0)$, $V = \frac{1}{2}(\partial^2/\partial\sigma\partial\sigma') F(\sigma_0, \sigma_0)$ and $\|\varepsilon_n\| = o(\|\sigma_n - \sigma_0\|)$. Furthermore, since $\|\sigma_n - \sigma_0\|$ is $O(n^{-1/2})$ we have that $\|\varepsilon_n\|$ is $o(n^{-1/2})$. This together with (3) implies that $n^{1/2}\Delta'V(\sigma_n - \sigma_0)$ tends to zero as $n \to \infty$. Therefore if $n^{1/2}(\sigma_n - \sigma_0)$ converges to a vector δ as in (2), then

$$\delta'V\Delta = 0. \qquad (5)$$

Equation (5) is a direct consequence of assumption (3). It shown that population drift ought to be chosen in such a way that the corresponding vector δ will be orthogonal to the column space of Δ with respect to the weight matrix V.

As in Proposition 1 it follows that $n^{1/2}(\hat{\theta} - \theta_0)$ is asymptotically normal with mean vector $\mathbf{0}$ and covariance matrix

$$\Pi = (\Delta'V\Delta)^{-1}\Delta'V\Gamma V\Delta(\Delta'V\Delta)^{-1}, \qquad (6)$$

so that $\hat{\theta}$ remains asymptotically unbiased. If $V = \Gamma^{-1}$, then (6) simplifies to

$$\Pi = (\Delta'\Gamma^{-1}\Delta)^{-1}.$$

Notice that assumption (3), which leads to an asymptotically unbiased estimator $\hat{\theta}$, requires that the population drift vector δ employed in (2) be defined *with respect to the particular discrepancy function chosen* (through the matrix V in (5)).

Discrepancy functions satisfying the condition $V = \Gamma^{-1}$ are called *correctly specified* (Browne, 1984). For example, in the analysis of covariance structures the discrepancy functions F_{ML} and F_{GLS} are correctly specified if the sample is drawn from a normally distributed population. A minor modification to assumption (3) may be made when the class of correctly specified discrepancy functions is under consideration. This is

$$\|\theta_{0,n} - \theta_0\| = o(n^{-1/2}). \qquad (7)$$

Then the argument leading to (5) still applies. Moreover (7) and (4) imply that θ_0 is completely determined by the matrix V. Therefore *for the class of correctly specified discrepancy functions the population value θ_0 is uniquely defined by condition (7) and the estimator $\hat{\theta}$ is an asymptotically unbiased estimator of θ_0.*

References

Bentler, P M (1985), *Theory and Implementation of EQS A Structural Equations Program* (BMDP Statistical Software, Los Angeles).

Bentler, P.M. and T Dijkstra (1985), Efficient estimation via linearization in structural models, in P. Krishnaiah, ed, *Multivariate Analysis* - VI (Elsevier Science Publishers, Amsterdam) pp 9-42

Browne, M W (1982), Covariance structures, in D M Hawkins, ed, *Topics in Applied Multivariate Analysis* (Cambridge University Press, Cambridge) pp 72-141

Browne, M W (1984), Asymptotically distribution-free methods for the analysis of covariance structures, *Br J Math Statist Psychol* **37**, 62-83

Dijkstra, T (1983), Some comments on maximum likelihood and partial least squares methods, *J Econometrics* **22** 67-90

Foutz, R V. and R C Srivastava (1977), The performance of the likelihood ratio test when the model is incorrect, *Ann Statist* **5**, 1183-1194

Huber, P J (1967), The behavior of maximum likelihood estimates under nonstandard conditions, in (ed. J. Neyman), *Proc 5th Berkeley Symp Math Statist and Prob*, Vol 1 (University of California Press, Berkeley) pp. 221–333

Shapiro, A (1983), Asymptotic distribution theory in the analysis of covariance structures (a unified approach), *S Afr Statist J* **17** 33–81.

Shapiro, A (1985), Asymptotic equivalence of minimum discrepancy estimators to GLS estimators, *S Afr Statist J* **19**, 73–81

Stroud, T.W. (1972), Fixed alternatives and Wald's formulation of the noncentral asymptotic behavior of the likelihood ratio statistic, *Ann Math Statist* **43**, 447–454

Wald, H (1943), Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Trans Amer Math Soc* **54** 426–482

White, H. (1982), Maximum likelihood estimation of misspecified models, *Econometrica* **50**, 1–25