STRUCTURAL EQUATION MODELING
2026, VOL. 33, NO. 1, 32–38
ISSN: 1070-5511 | E-ISSN: 1532-8007



# Fitting Covariance Structure Models With Arbitrary Observed Distributions: Strong Convergence of the Asymptotically Distribution-Free Estimator

Tenko Raykov[a], Spiridon Penev[b], and Bingsheng Zhang[c]

[a]Michigan State University; [b]University of New South Wales, Sydney; [c]StataCorp LLC

**ABSTRACT**
The large-sample behavior of the asymptotically distribution-free estimator in covariance structure analysis is studied. It is shown that with increasing sample size and under suitable conditions, the estimator almost surely converges to the true parameter value. This strong convergence (i) represents numerical convergence with probability 1 of the resulting estimates of model parameters to their respective population values, as well as (ii) is stronger than the consistency and convergence in distribution of a parameter estimator, thus implying the latter two types of large-sample behavior. The demonstrated asymptotic convergence of the asymptotically distribution-free estimator for covariance structure models is illustrated on data.

**KEYWORDS**
Asymptotically distribution-free estimator; consistency; convergence in distribution; covariance structure analysis; discrepancy function; strong convergence

Covariance structure analysis (CSA) has become increasingly popular in the behavioral and social sciences over the past half century or so (e.g., Klein, 2023). A main reason for its widespread use in these and cognate disciplines is the fact that CSA permits improved estimation of the parameters of theoretically and empirically relevant models of studied phenomena while accounting for measurement error in the explanatory and outcome variables (e.g., Bollen, 1989). One beneficial feature of the resulting estimates is their consistency as well as large-sample normality and efficiency under certain conditions (e.g., Browne & Arminger, 1995). This asymptotic estimator behavior responds to a consequential query for behavioral and social scholars, which asks about possible trends underlying the estimates of considered model parameters as sample size increases without bound (cf. Raykov, 2019).

A highly desirable feature of a parameter estimator in CSA as well as other statistics applications is the diminishing distance, with growing amounts of sample information, between the estimates rendered by that estimator and the pertinent population parameter value. For this reason, the large-sample behavior of popular estimators, like those provided by maximum likelihood or Bayesian inference methods for instance, has been the concern of a large body of statistical, methodological, and related literature (cf. DasGupta, 2008). Most of the time, however, what was meant in respective publications when referring to the behavior of these and other estimators with increasing sample size, is their consistency as well as their convergence in distribution, in particular for their applications in CSA (cf. Browne, 1984, and references therein).

The present article intends to complement the cited and related research on the large-sample behavior of parameter estimators for covariance structure models. To this end, we discuss the asymptotic behavior of a more general estimator, the asymptotically distribution-free (ADF) estimator by Browne (1984). This estimator does not in effect make assumptions about the observed variable distributions, and does not need to use prior distributions in order to proceed. In addition, we are concerned with a stronger convergence property than consistency and convergence in distribution of parameter estimators. Specifically, we demonstrate below the strong convergence with increasing sample size of the ADF estimator of covariance structure model parameters (this convergence feature is alternatively and interchangeably also referred to as strong consistency, almost sure convergence, or convergence with probability 1; e.g., Ferguson, 1996). There are two reasons for our interest in the strong convergence of parameter estimators, in particular in CSA. One is that this convergence implies both consistency and convergence in distribution, while the converse is not true—that is, strong convergence does not follow from consistency and/or convergence in distribution (e.g., Billingsley, 2011). The other reason is the observation that in our experience it is usually strong convergence that many empirical behavioral and social scientists tend to mean or even assume, at times intuitively, when referring to either consistency or convergence in distribution for estimators of model parameters under consideration.

The plan of the article is as follows. We commence with a brief non-technical discussion of large-sample properties of sequences of random variables, such as parameter estimators, with increasing sample size. We then demonstrate the almost sure convergence of the ADF estimator to the corresponding population parameter value for covariance structure models under suitable conditions. Subsequently, we illustrate the relevance and utility of this asymptotic property for empirical research by employing data sets with increasing sample sizes.

## 1. Large-Sample Behavior of Parameter Estimators

In this section, for the aims of the present note we briefly discuss the three large-sample behaviors that relate to the following discussion (for a more detailed, non-technical discussion we refer, e.g., to Raykov, 2019, and references therein). To this end, we begin with the observation that studying the large-sample behavior of an estimator for a parameter of interest - which is a vector with at least one and generally two or more components - involves (a) considering the sequence of random variables that represents this estimator at increasing sample sizes, and (b) examining the behavior of this sequence (e.g., Ferguson, 1996). Three types or modes of such large-sample behavior are of importance to this article, and we describe them in turn next.

Perhaps the most popular and frequently used type of estimator convergence when sample size grows unboundedly, is that of convergence in distribution (also called convergence in law, or weak convergence). A widely used example of such convergence is provided by the central limit theorem (e.g., Wasserman, 2004), which states that the (rescaled) sampling distribution of the mean approaches under fairly mild conditions a standard normal distribution (when the mean is appropriately normed). This is the weakest asymptotic convergence type of all three modes mentioned (hence its name as well), in that it follows from each of the other two modes of convergence. Practically speaking, weak convergence means that the distribution of the parameter estimator approximates with growing sample size increasingly well that limiting distribution (i.e., the standard normal in the last example; cf. Raykov et al., 2022). In other words, and with some simplification, when convergence in distribution holds then with increasing sample size the cumulative distribution function (CDF) of the estimator in question becomes more and more like the limiting distribution's CDF, and the difference between these two CDFs gradually in effect disappears (at any point of continuity of the limiting distribution, and with very large samples if need be). This convergence type is weaker than almost sure convergence, since the latter implies both consistency and convergence in distribution but the converse is not true, as pointed out earlier.

The next in popularity large-sample behavior of parameter estimators is the convergence in probability, which is often referred to as consistency and is also highly desirable for a parameter estimator. Accordingly, as sample size increases, the probability of the population parameter value being close (very close) to its estimator grows as well and approaches 1. In other words, with increasing sample size the probability of finding the estimator outside of a prespecified (short) range around that parameter value, becomes smaller and smaller, and eventually in effect vanishes (with very large samples if need be). However, as shown previously (e.g., DasGupta, 2008), neither convergence in distribution nor consistency guarantees that as sample size increases the resulting sample parameter estimates themselves, as real numbers (in general, as vectors of real numbers), become closer and closer to that population value (in a calculus sense; e.g., Apostol, 2006). This latter feature will hold, with probability 1, only if the estimator sequence converges almost surely to that true value. In the latter case, the sequence of numbers representing the estimator value at increasing sample size, converge as real numbers and in a calculus sense to that true value.

As can be observed readily, the last is a very strong statement. In fact, as indicated, strong convergence is the strongest of all three convergence types mentioned in this section. This means that once almost sure convergence holds then also (a) convergence in probability, and (b) convergence in distribution, are the case with increasing sample size. The fact that the latter two convergence types (modes) follow from strong convergence, is the reason why this article is interested in strong convergence with respect to the ADF estimator in CSA. Specifically, in the next section we show that when sample size grows unboundedly, the ADF estimator in CSA almost surely (i.e., with probability 1) converges to the population/true parameter value for a considered covariance structure model, under a set of regularity conditions that can be seen as relatively mildly restrictive.

## 2. Strong Convergence of the Asymptotically Distribution-Free Estimator in Covariance Structure Analysis

### 2.1. Background, Notation, and Assumptions

As mentioned in the introductory section, this note builds on Browne (1984) and complements his findings with the strong convergence result of the ADF estimator in CSA, as stated in Proposition 1 below. To accomplish this goal, we start by assuming that a covariance structure model, denoted M, is of interest for a set of $p$ observed variables that are symbolized by the $p$ x 1 vector $\underline{y}$ ($p > 1$; underlining denotes vector and priming transposition in this article). We presume in the remainder that M is (a) a special case of the highly popular and widely used, general LISREL model in CSA (e.g., Jöreskog & Sörbom, 1996), and (b) considered for fitting to an available data set collected for these variables on a sample of $n$ studied independent units of analysis from a population under investigation ($n > 1$). The model is stipulated as identified throughout this note, as well as containing at least one free (unconstrained) parameter.[1] We denote by $\underline{\gamma}$ the vector of all parameters of M, and by $\Gamma$ the pertinent model parameter space. We designate the

[1]For models with two or more factor loadings, without limitation of generality it is presumed throughout this note that they are positive in the population (which can be usually ensured by indicator reversal, e.g., reverse scoring of response options, if needed).

population covariance matrix of $\underline{y}$ by $\Sigma_0$, its empirical counterpart generically by S, and assume $\Sigma_0$ to be positive definite (e.g., Johnson & Wichern, 2018).

For our purposes in what follows, we will denote by $\underline{\sigma}$ the strung-out vector of the non-redundant elements of a generic population covariance matrix $\Sigma$, and by $\underline{s}$ that vector for S. As is well known, a covariance structure analysis model is reflected in its implied covariance matrix $\Sigma(\underline{\gamma})$ (e.g., Browne & Arminger, 1995), whose strung-out version is designated as $\underline{\sigma}(\underline{\gamma})$ in the sequel. With this notation, the model can be generally represented by the matrix equation $\Sigma = \Sigma(\underline{\gamma})$ (that is, $\underline{\sigma} = \underline{\sigma}(\underline{\gamma})$). The last equation (a) states that the model fits perfectly the population covariance matrix $\Sigma$, and (b) can be considered the null hypothesis in CSA with respect to the model in question (cf. Bollen, 1989). In addition, we will denote by $\Delta(\underline{\gamma})$ the Jacobian of $\underline{\sigma}(\underline{\gamma})$, which contains the first derivatives of $\underline{\sigma}(\underline{\gamma})$ with respect to $\underline{\gamma}$ (i.e., $\Delta(\underline{\gamma}) = \frac{\partial \underline{\sigma}}{\partial \underline{\gamma}'}$; e.g., Apostol, 2006). As is commonly the case in CSA, we will refer to a model as correct (true, valid, or holding), if there exists a point $\underline{\gamma}_0$ in $\Gamma$, such that $\Sigma_0 = \Sigma(\underline{\gamma}_0)$ (i.e., $\underline{\sigma}_0 = \underline{\sigma}(\underline{\gamma}_0)$) is true.

Furthermore, given an empirical covariance matrix, denoted S as above, for the aims of this article we will be interested in obtaining an estimate $\underline{\gamma}^*$ of $\underline{\gamma}$ by minimizing over $\Gamma$ a discrepancy function of S and $\Sigma$, which will be denoted by $F(S, \Sigma)$. This function is assumed in the sequel to be twice continuously differentiable, non-negative, and possessing the value of 0 only when $S = \Sigma$ (Browne, 1984, p. 64). Specifically, for a particular covariance structure model, the function $F(S, \Sigma) = F(S, \Sigma(\underline{\gamma}))$ attains its minimum at $\underline{\gamma}^*$ in $\Gamma$, and we will denote by $\Sigma^* = \Sigma(\underline{\gamma}^*)$ the reproduced covariance matrix by the model at that point, $\underline{\gamma}^*$. Accordingly, the rest of this note is concerned with quadratic discrepancy functions of the type

$$F(S, \Sigma(\underline{\gamma}) \mid W) = (\underline{s} - \underline{\sigma}(\underline{\gamma}))' \; W^{-1}(\underline{s} - \underline{\sigma}(\underline{\gamma})), \tag{1}$$

where W is a generic symmetric weight matrix of size $p(p+1)/2 \times p(p+1)/2$, which is assumed to be positive definite (denoted $W > 0$ in the sequel). As implied from Equation (1), this type of discrepancy functions evaluate the weighted distance between the sample covariance matrix S and the implied covariance matrix by a considered model for any given set of values for its parameters.

The developments in the remainder of this article make the following assumptions (Browne, 1984, pp. 64–66):

(A1) The covariance structure model under consideration, M, is identified and true;

(A2) All observed variables in $\underline{y}$ have finite eighth-order moments (and hence all their lower-order moments are finite as well; e.g., Arnold, 1990; see also Footnote 2);

(A3) The discrepancy function $F(\Sigma_0, \Sigma) = F(\Sigma_0, \Sigma(\underline{\gamma}))$ has a unique minimum over $\Gamma$, which is achieved at $\underline{\gamma} = \underline{\gamma}_0$ that is an interior point of $\Gamma$;

(A4) $\Delta_0 = \Delta(\underline{\gamma}_0)$ has full column rank denoted $q$;

(A5) The model parameter space $\Gamma$ is closed and bounded; and

(A6) $\Delta(\underline{\gamma})$ and $\Sigma(\underline{\gamma})$ are continuous functions of $\underline{\gamma}$ .[2]

Under the above assumptions, the proposition in the next subsection shows the strong convergence of the ADF estimator in CSA with increasing sample size. Due to our interest in this estimator, we will use the notation $\underline{\gamma}^*$ for it hereafter, in particular for this estimator of the parameters of a covariance structure model of interest. Also, we will attach the subindex $n$ to each object that is meant to be considered at a given sample size, $n$; for instance, $S_n$ and $\underline{\gamma}_n^*$ will denote correspondingly the covariance matrix and ADF estimator in a sample of size $n$ (and similarly for the other sub-indexed quantities appearing below).

## 2.2. Strong Convergence of the ADF Estimator in Covariance Structure Analysis

In this subsection, we will be concerned with the following main result of the present article.

**Proposition 1:** Under conditions (A1) through (A6), the ADF estimator in CSA, $\underline{\gamma}_n^*$, strongly converges to the true population parameter as sample size $n$ increases without limit.

Indeed, consider first the quantity $Q_n = (\underline{s}_n - \underline{\sigma}(\underline{\gamma}_0))' \; W_n^{-1} \; (\underline{s}_n - \underline{\sigma}(\underline{\gamma}_0))$, with $W_n$ being an estimator of the asymptotic covariance matrix of the elements of the sample covariance matrix of $\underline{y}$ at a given sample size, $n$. This estimator is assumed to be strongly converging with increasing sample size to a positive definite weight matrix, denoted for convenience W, with entries defined as in Browne (1984, p. 71, Equations (3.1) through (3.4); defining the entries of $W_n$ as in the last cited 3 equations in terms of the participating empirical second and fourth-order moments in them, satisfies this assumption). That is, this quantity $Q_n$ is a quadratic form at a given sample size $n$, which we denote $F_{n,0}$ for notational clarity in the sequel, i.e., formally we set

$$F_{n,0} = (\underline{s}_n - \underline{\sigma}(\underline{\gamma}_0))' \; W_n^{-1}(\underline{s}_n - \underline{\sigma}(\underline{\gamma}_0)). \tag{2}$$

We notice that $F_{n,0}$ represents the above discrepancy function value for given (a) weight matrix that is assumed positive definite (at least with probability 1), (b) model, (c) empirical covariance matrix, (d) population parameter value, and (d) sample size (see Equation (1)).

We next observe the following facts: (i) the elements of $\underline{s}_n$ strongly converge to their corresponding elements of $\underline{\sigma}_0$, as do the elements of $W_n$ to the respective elements of its population counterpart W (cf. Billingsley, 2011; see also Footnote 3); (ii) $\underline{\sigma}(\underline{\gamma}_0) = \underline{\sigma}_0$ due to the model under

[2]Assumption (A2), which stipulates finite eighth-order moments, will be satisfied in the vast majority - if not all - contemporary empirical behavioral and social science research settings, due to this research working exclusively with individual realizations of random variables that are bounded both from below and above (e.g., Arnold, 1990). This assumption is included in the note in order to ensure more general applicability of its finding in Proposition 1 (e.g., in future studies where a random variable of concern need not be bounded in this way).

consideration, M, being correct; and (iii) the right-hand side of Equation (2) is a non-negative continuous function in each element of the matrices $\underline{s}_n$, $\underline{\sigma}(\underline{\gamma}_0)$, and $W_n$.[3]

Then, due to the last observation (iii), from the continuity mapping theorem (e.g., Billingsley, 2011; see also Raykov, 2019) it follows that

$$F_{n,0} \rightarrow_{a.s.} 0, \tag{3}$$

where $\rightarrow_{a.s.}$ denotes almost sure convergence (strong convergence) as $n \rightarrow \infty$ (with the last notation meaning sample size increasing without limit).

If we next designate by $\underline{\gamma}_n^*$ the ADF estimator at sample size $n$ (as indicated above) that results by minimizing the discrepancy function (1) for a sample of size $n$ and a positive definite weight matrix, we observe that $\underline{\gamma}_n^*$ minimizes the following function of $\underline{\gamma}$, denoted $F_n(\underline{\gamma})$, over $\Gamma$ :

$$F_n(\underline{\gamma}) = (\underline{s}_n - \underline{\sigma}(\underline{\gamma}))' \; W_n^{-1}(\underline{s}_n - \underline{\sigma}(\underline{\gamma})). \tag{4}$$

Denote now by $F_n{}^*$ the value of $F_n$ at $\underline{\gamma} = \underline{\gamma}_n^*$, that is,

$$F_n{}^* = (\underline{s}_n - \underline{\sigma}(\underline{\gamma}_n^*))' \; W_n^{-1}(\underline{s}_n - \underline{\sigma}(\underline{\gamma}_n^*)). \tag{5}$$

Owing to $\underline{\gamma}_n^*$ being the minimizer of the right-hand side of Equation (4), i.e., the function $F_n(\underline{\gamma})$, over the model parameter space $\Gamma$, while $\underline{\gamma}_0$ is a point in $\Gamma$, it follows that

$$F_n{}^* \leq F_{n,0}. \tag{6}$$

Then from (a) Inequality (6), (b) the fact that $F_n{}^* \geq 0$ since (5) is a positive definite quadratic form (with probability 1, due to $W_n > 0$ at least with probability 1), and (c) the strong convergence of $F_{n,0}$ to 0 as shown above (see Equation (3)), it follows that also

$$F_n{}^* \rightarrow_{a.s.} 0 \tag{7}$$

as $n \rightarrow \infty$.

We next observe the following fact of general validity. Since any symmetric positive definite matrix, like W in Equation (1) defining the generic ADF discrepancy function, can be expressed as the square of its associated square-root matrix that is also symmetric and positive definite (e.g., Johnson & Wichern, 2018), the right-hand side of Equation (1) can be rewritten as

$$\begin{aligned} &(\underline{s} - \underline{\sigma}(\underline{\gamma}))' \; W^{-1}(\underline{s} - \underline{\sigma}(\underline{\gamma})) \\ &= (\underline{s} - \underline{\sigma}(\underline{\gamma}))' \; W^{-1/2}(W^{-1/2})' \; (\underline{s} - \underline{\sigma}(\underline{\gamma})) \\ &= \left[(W^{-1/2})' \; (\underline{s} - \underline{\sigma}(\underline{\gamma}))\right]' \left[(W^{-1/2})' \; (\underline{s} - \underline{\sigma}(\underline{\gamma}))\right]. \end{aligned} \tag{8}$$

The last presented, generic argument holds also for (i) the empirical (sample) covariance matrix, $S_n$, at a given sample of size $n$; and (ii) the weight matrix $W_n$ employed for ADF estimation purposes then (due to the latter matrix being positive definite with probability 1; see above). Therefore, it follows that the discrepancy function value $F_n{}^*$ in Equation (5), is representable as the product $\underline{x}_n{}' \; \underline{x}_n$, i.e.,

$$F_n{}^* = \underline{x}_n{}'\underline{x}_n \tag{9}$$

holds, where

$$\underline{x}_n = \left[(W_n{}^{-1/2})' \; (\underline{s}_n - \underline{\sigma}(\underline{\gamma}_n^*))\right]. \tag{10}$$

Based on Equations (9) and (10), the strong convergence of $F_n{}^*$ to 0 shown above (see Equation (7)) along with its non-negativity, implies this convergence of $\underline{x}_n$ as $n$ increases without bound, that is,

$$W_n{}^{-1/2}(\underline{s}_n - \underline{\sigma}(\underline{\gamma}_n^*)) \rightarrow_{a.s.} 0 \tag{11}$$

holds as $n \rightarrow \infty$. Due to $W_n{}^{-1/2}$ being invertible (with probability 1) as a positive definite matrix, the convergence stated in (11) implies the strong convergence of $\underline{s}_n - \underline{\sigma}(\underline{\gamma}^*{}_n)$ itself, that is,

$$\underline{s}_n - \underline{\sigma}(\underline{\gamma}_n^*) \rightarrow_{a.s.} 0 \tag{12}$$

as $n \rightarrow \infty$.

As a last step, we recall that the covariance structure model under consideration was assumed to be correct (i.e., true; see assumptions (A1) through (A6)). This implies, as indicated earlier, that

$$\underline{s}_n - \underline{\sigma}(\underline{\gamma}_0) \rightarrow_{a.s.} 0 \tag{13}$$

when $n \rightarrow \infty$, that is,

$$\underline{s}_n \rightarrow_{a.s.} \underline{\sigma}(\underline{\gamma}_0) \tag{14}$$

as $n \rightarrow \infty$. Then it follows from (12) and (13) that

$$\underline{\sigma}(\underline{\gamma}_n^*) \rightarrow_{a.s.} \underline{\sigma}(\underline{\gamma}_0) \tag{15}$$

for $n \rightarrow \infty$.

We recall now the earlier assumption of the model being identified, which means that each of its parameters is expressible in at least one but up to finitely many ways as a continuous function of the elements of the population covariance matrix (e.g., Bollen, 1989). Based on this last continuity observation, employing the earlier utilized continuity mapping theorem again, one concludes finally that

$$\underline{\gamma}_n^* \rightarrow_{a.s.} \underline{\gamma}_0 \tag{16}$$

as $n \rightarrow \infty$.

Relationship (16) states that the ADF parameter estimator in CSA strongly converges to the true parameter population value as sample size increases without bound, which statement was of focal interest in the present section. In the next section, we illustrate this behavior of the ADF estimator using data sets with increasing size.

## 3. Illustration on Data

We exemplify here the result of Proposition 1 by employing a confirmatory factor analysis (CFA) setting that can be of special relevance in behavioral and social measurement research (cf., e.g., Raykov & Marcoulides, 2011). To this

[3]The elements of the weight matrix W referred to in this section are the same functions of the population second- and fourth-order moments of the observed variables in $\underline{y}$, as the functions representing the elements of $W_n$ in terms of the sample second- and fourth-order moments (Browne, 1984, p. 71, Equations (3.1) through (3.4)). Specifically, the elements of $W_n$ are $w_{ijkl}$ - $w_{ij}$ $w_{kl}$, where $w_{ijkl}$ denotes the empirical fourth-order mixed central moment for the quadruple ($y_i$, $y_j$, $y_k$, $y_l$) while $w_{ij}$ symbolizes (generically) the second-order mixed central moment for $y_i$ and $y_j$, for the $i$th, $j$th, $l$th, and $k$th elements of the observed variable vector $\underline{y}$ ($i, j, k, l = 1, \ldots, p$).

end, we utilize a sequence of $r = 19$ simulated data sets for successive sample sizes of $n = 500$, 1000, 2000, 5000, 10,000, 20,000, 50,000, 100,000, 200,000, 500,000, 1,000,000, 1,500,000, 2,000,000, 2,500,000, 5,000,000, 10,000,000, 20,000,000, 50,000,000, and 100,000,000 independent cases. These $r$ data sets were generated using the two-factor model defined by Equation (17) for $k = 6$ observed variables $y_1$, …, $y_6$ (cf. Mulaik, 2009). At each of these 19 sample sizes, the observed outcome variables were simulated from distinct symmetric or asymmetric distributions that were almost all non-normal distributions. The data generation model used thereby was defined as follows:

$$\begin{aligned} y_1 &= \mu_1 + \lambda_1 f_1 + e_1, \\ y_2 &= \mu_2 + \lambda_2 f_1 + e_2, \\ y_3 &= \mu_3 + \lambda_3 f_1 + e_3, \\ y_4 &= \mu_4 + \lambda_4 f_2 + e_4, \\ y_5 &= \mu_5 + \lambda_5 f_2 + e_5, \\ y_6 &= \mu_6 + \lambda_6 f_2 + e_6, \end{aligned} \tag{17}$$

where $f_1$ and $f_2$ were standard normal variates with correlation $\rho = Corr(f_1, f_2) = .4$ (with $Corr(.)$ denoting correlation), while $\mu_1 = \mu_2 = \ldots = \mu_6 = 0$ and $\lambda_1 = 2.0$, $\lambda_2 = 1.3$, $\lambda_3 = 1.5$, $\lambda_4 = 1.2$, $\lambda_5 = 2.1$, and $\lambda_6 = 1.4$ were set. The error terms $e_1$, …, $e_6$ were thereby simulated as adhering to a centered normal, chi-square(2), exponential, log-normal, chi-square(5), and gamma distribution, respectively. (The Appendix provides the Stata code employed for simulating all datasets and fitting the two-factor model using ADF estimation, and states the specific details regarding the last-mentioned six distributions and their parameters, including the used chi-square distributions with 2 and 5 degrees of freedom, respectively.) The resulting estimates of the parameter that is usually of main interest in a CFA setting, the latent correlation $\rho$, are presented in Table 1.[4,5]

As seen from Table 1, the ADF estimates stabilize at the population value of $\rho = .400$ beginning with sample size 10,000,000, and remain at that value for the subsequent larger samples (see also Note to Table 1; cf. Raykov, 2019). This stabilization behavior is consistent with the statement of Proposition 1, which asserts the strong convergence of the ADF estimator with increasing sample size, meaning numerical convergence with probability 1 as sample size increases. In particular, the correlation estimates' behavior observed in Table 1 is logically implied by that generally valid proposition (under its assumptions that can be treated as fulfilled here) in the present CFA special case used for its illustration.

**Table 1.** Asymptotically distribution-free estimates of the latent factor correlation, $\rho$, at increasing sample sizes in the fitted two-factor model used in the illustration section.

| Sample size ($n$) | $\rho$ | Correlation estimate ($\rho_n^*$) |
|---|---|---|
| 500 | .400 | .353 |
| 1,000 | .400 | .424 |
| 2,000 | .400 | .402 |
| 5,000 | .400 | .415 |
| 10,000 | .400 | .397 |
| 20,000 | .400 | .397 |
| 50,000 | .400 | .397 |
| 100,000 | .400 | .399 |
| 200,000 | .400 | .402 |
| 500,000 | .400 | .400 |
| 1,000,000 | .400 | .398 |
| 1,500,000 | .400 | .400 |
| 2,000,000 | .400 | .401 |
| 2,500,000 | .400 | .401 |
| 5,000,000 | .400 | .399 |
| 10,000,000 | .400 | .400 |
| 20,000,000 | .400 | .400 |
| 50,000,000 | .400 | .400 |
| 100,000,000 | .400 | .400 |

*Notes:* $\rho$ = population latent correlation ($\rho = .400$; see Equation (17) and their immediately following discussion); $\rho_n^*$ = ADF estimate of the latent correlation, at corresponding sample size, $n$ (see also Footnote 5). One might argue that with increasing sample size the ADF estimates of the latent correlation in this Table 1 appear to converge (also) in probability to the true correlation value of .4. We may emphasize in response that convergence in probability follows from almost sure convergence, but not reversely (for the latter, see Arnold, 1990). Since almost sure convergence holds in the present example (due to Proposition 1 being a general statement while the example is a special case of it), it follows that also convergence in probability holds in this example.

## 4. Conclusion

This note was concerned with the almost sure, or with probability 1, convergence of the asymptotically distribution-free estimator (Browne, 1984) in covariance structure analysis (strong convergence or strong consistency). Proposition 1 in the last section, under its explicated conditions and assumptions (A1) through (A6), demonstrated this convergence for each model parameter in a considered covariance structure model to its counterpart population parameter value. This almost sure consistency property is a stronger statement than (a) the ADF estimator convergence in distribution, and (b) its convergence in probability or consistency (see Browne, 1984, and references therein). The reason is the fact that strong convergence implies both (a) and (b), but the converse is not true. This convergence with probability 1 finding of the present article solidifies what may be referred to as intuitive expectation by many behavioral and social scientists, that with the increase of study size (sample size) the resulting numerical sequence of ADF estimates practically always approach infinitesimally (converge, as real numbers, on) their corresponding parameter population values for a covariance structure model under consideration. We emphasize that for this convergence statement to hold, no distributional assumptions need to be made with respect to the observed measures (manifest or observed variables) that participate in the model (see Footnote 2 for the empirical validity of the assumption of finite eight-order moments in contemporary behavioral and social research).

Several limitations of the discussion in this note need to be pointed out. One is the fact that the article is not concerned with speed of convergence. More specifically, neither Proposition 1 nor any part of this note indicate (or are meant

[4]All simulated data sets were associated with no error messages during the pertinent data generation and model fitting process.

[5]Since Proposition 1 is generally valid (under its conditions that can be treated as holding in this section of the main text), the choice of rounding-off to third decimal digit in Table 1 is not essential for the illustration aims with the special case used in the present section.

to indicate) at what sample size the obtained ADF estimates can be considered practically equal (or within a prespecified distance) from the true population values of the model parameters. Similarly, the article does not have any direct implications as to what sample sizes may lead to numerical issues during ADF estimation, e.g., due to possibly insufficient sample size considering also the need to estimate the potentially large-sized weight matrix necessary for this estimation procedure (see Equations (1) and (2) as well as their surrounding discussion). Given the complexity of these issues, we encourage future studies on speed of this almost sure convergence, including simulation studies that go beyond the confines of this note. Relatedly (and as stated in its title and pointed out earlier), the concern of the note is merely with demonstrating the almost sure convergence of the ADF estimator (under the conditions pointed out above). With this in mind, we are not implying a recommendation for wider use of this estimator in its original form (see Equation (2)) at less than "optimal sample sizes." This is due to the well-known fact that use of the estimator necessitates estimation of a rather large number of parameters, depending also on model complexity and number of observed variables. Sample sizes that are supportive of that ADF estimator utilization can therefore be very high in certain circumstances, perhaps well into the thousands, according to considerable simulation research over the past several decades (e.g., Hu et al., 1992). At the same time, however, main beneficial features of ADF estimation can be capitalized on in empirical behavioral and social research in some settings with smaller sample sizes by using modified ADF estimators, which deal with the potential ill-conditioning of needed large-scale covariance matrices and instability in estimation of higher-order moments (e.g., Chun et al., 2018; Foldnes et al., 2019; Muthén & Muthén, 2025; see also Foldnes et al., 2025). Some of them have been already made widely and routinely available in popular software, and could be recommended to consider using with such samples (e.g., the Mplus' weighted least squares with mean and variance adjustment estimator; Muthén & Muthén, 2025).

Second, as can be implied from the pertinent calculus-based argument (cf. Apostol, 2006), Proposition 1 does not imply that once reaching a particular (pre-specified) "minimal" proximity to the respective population value, the ADF parameter estimates after that sample size will all remain within that vicinity (or stay essentially identical to the population parameter values; cf. Raykov et al., 2022, and Table 1). The reason is that the proposition is only of asymptotic nature, which makes it impossible to come up with more concrete, finite-sample interpretations of this type. Therefore, with Proposition 1 still holding, as sample size increases without bound the parameter estimates could "step out" of, and then "return back" into a pre-specified vicinity of the population parameter value, a finite number of times.

Third, as pointed out in prior publications on strong convergence, Proposition 1 per se does not ensure that in every empirical study or setting the ADF estimates with increasing sample size will converge as a numerical sequence on the corresponding true parameter population value (cf. Raykov, 2019). The reason is that the actual statement of the proposition is only convergence that holds apart from an event with zero probability, which does not exclude the possibility of such an event occurring in an application. Moreover, we assumed throughout the article that the model of interest was true to begin with. In empirical research, however, this is rarely if ever the case. Hence, it may be conjectured that with a sufficiently small model misspecification (cf. Browne, 1984), the ADF estimates with increasing sample size could become sufficiently close to the population parameter value (with "sufficiently" depending also on model, parameter, and sample size). This conjecture needs to be examined in detail, however, which is outside the frame of the present note.

Last but not least, Proposition 1 is based on a set of regularity conditions stated earlier (see (A1) through (A6) in the last subsection and their preceding discussion, as well as Footnote 2). Thereby, any of them alone, or an incomplete combination of them, need not imply the ADF estimator's strong convergence. Moreover, none of these conditions, nor all of them or in combination, contains sufficient information that would allow a scholar to determine a minimal sample size beyond which the resulting ADF parameter estimates will fall within a pre-specified neighborhood of the population value of a given parameter in a covariance structure model of interest.

In conclusion, this article demonstrated the strong convergence with increasing sample size of the most general and widely applicable parameter estimator in covariance structure analysis, the asymptotically distribution-free estimator (Browne, 1984), under suitable, relatively mild conditions. The note showed that with growing sample size and regardless of the observed variable distribution, this estimator furnishes estimates that with probability 1 converge in a calculus sense as real numbers on the population values of the model parameters for a considered covariance structure model (see also Footnote 2). This large-sample behavior is arguably assumed or perhaps even expected by many empirical behavioral and social scientists, but does not follow from consistency and/or convergence in distribution shown in earlier research. Instead, this ADF estimator convergence with probability 1 is ensured by Proposition 1 of the present article, under its above-mentioned assumptions.

## Author Note

The contributions of the three authors were equivalent. We are indebted to the Editor and two anonymous Referees for valuable comments and criticism on an earlier version, which have contributed considerably to its improvement.

## References


Apostol, T. (2006). *Calculus*. Wiley.

Arnold, S. F. (1990). *Mathematical statistics*. Prentice-Hall.

Billingsley, P. (2011). *Probability and measure*. Wiley.

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *The British Journal of Mathematical and Statistical Psychology*, *37*(Pt 1), 62–83. https://doi.org/10.1111/j.2044-8317.1984.tb00789.x

Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185–249). Springer.

Chun, S. Y., Browne, M. W., & Shapiro, A. (2018). Modified distribution-free goodness-of-fit test statistic. *Psychometrika*, *83*, 48–66. https://doi.org/10.1007/s11336-017-9574-9

DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. Springer.

Ferguson, T. S. (1996). *A course in large sample theory*. Chapman & Hall.

Foldnes, N., Marcoulides, G. A., & Ulf Olsson, U. H. (2019). Examining the performance of the modified ADF goodness-of-fit test statistic in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*, 778–789. https://doi.org/10.1080/10705511.2019.1586545

Foldnes, N., Moss, J., & Grønneberg, S. (2025). Improved goodness of fit procedures for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *32*, 1–13. https://doi.org/10.1080/10705511.2024.2372028

Hu, L.-T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*, 351–362. https://doi.org/10.1037/0033-2909.112.2.351

Johnson, R. A., & Wichern, D. W. (2018). *Applied multivariate statistical analysis*. Pearson.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL8 user reference guide*. Scientific Software.

Klein, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford.

Mulaik, S. A. (2009). *Factor analysis*. Chapman-Hall.

Muthén, L. K., & Muthén, B. O. (2025). *Mplus user's guide*. Muthén & Muthén.

Raykov, T. (2019). Strong convergence of the coefficient alpha estimator for reliability of multiple-component measuring instruments. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*, 430–436. https://doi.org/10.1080/10705511.2018.1515019

Raykov, T., Doebler, P., & Marcoulides, G. A. (2022). Applications of Bayesian confirmatory factor analysis in behavioral measurement: Strong convergence of a Bayesian parameter estimator. *Measurement*, *20*, 215–227.

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Taylor & Francis.

Wasserman, L. (2004). *All of statistics*. Springer.


## Appendix

### Stata Code for Data Simulation and Model Fitting Used in the Illustration Section

```
cscript
matrix Rho = J(19,3,.)
matrix colnames Rho = sample_size rho rho_adf
local i = 1
//simulate data sets used in example (set numerically the sample sizes)
foreach n in 500 1000 2000 5000 10,000 20,000 50,000///
        100,000 200,000 500,000 1,000,000 1,500,000 2,000,000///
        2,500,000 5,000,000 10,000,000 20,000,000 50,000,000 100,000,000
{
clear
//set number of observations (invoke the sample sizes)
set obs `n'
//set model parameters (set the intercepts, slopes, and factor correlation)
scalar mu1 = 0
scalar mu2 = 0
scalar mu3 = 0
scalar mu4 = 0
scalar mu5 = 0
scalar mu6 = 0
scalar lambda1 = 2.0
scalar lambda2 = 1.3
scalar lambda3 = 1.5
scalar lambda4 = 1.2
scalar lambda5 = 2.1
scalar lambda6 = 1.4
scalar rho = .4
matrix C = (1, rho\rho, 1)
set seed 19
//draw correlated bivariate latent variables (simulate f1 and f2)
drawnorm f1 f2, double corr(C)
//generate outcome variables (simulate the y's)
ge double y1 = mu1 + lambda1*f1 + rnormal(0,1)/sqrt(2)
ge double y2 = mu2 + lambda2*f1+(rchi2(2)-2)/sqrt(8)
ge double y3 = mu3 + lambda3*f1+(rexponential(2)-2)/sqrt(8)
ge double y4 = mu4 + lambda4*f2+(exp(rnormal(0,.5))-exp(.25/2))
ge double y5 = mu5 + lambda5*f2+(rchi2(5)-5)/sqrt(25)
ge double y6 = mu6 + lambda6*f2+(rgamma(.5,1)-.5)
//set up tabular presentation (for sample size and latent correlation)
matrix Rho[`i',1] = `n'
matrix Rho[`i',2] = rho
//fit model using ADF estimation
sem (F1 -> y1 y2 y3)///
        (F2 -> y4 y5 y6),///
        variance(F1@1 F2@1) method(adf)
//populate table with results
matrix Rho[`i',3] = _b[/cov(F1,F2)]
local i = `i'+1
}
matrix list Rho
```

*Note*. Annotating comments added after '//' sign, explicating the activities following pertinent comment.