Deletion statistic accuracy in confirmatory factor models

Jennifer Koran¹ and Fathima Jaffari²

Abstract
Social science researchers now routinely use confirmatory factor models in scale development and validation studies. Methodologists have known for some time that the results of fitting a confirmatory factor model can be unduly influenced by one or a few cases in the data. However, there has been little development and use of case diagnostics for identifying influential cases with confirmatory factor models. A few case deletion statistics have been proposed to identify influential cases in confirmatory factor models. However, these statistics have not been systematically evaluated or compared for their accuracy. This study evaluated the accuracy of three case deletion statistics found in the R package influence.SEM. The accuracy of the case deletion statistics was also compared to Mahalanobis distance, which is commonly used to screen for unusual cases in multivariate applications. A statistical simulation was used to compare the accuracy of the statistics in identifying target cases generated from a model in which variables were uncorrelated. The results showed that Likelihood distance and generalized Cook’s distance detected the target cases more effectively than the Chi-square difference statistic. The accuracy of the Likelihood distance and generalized Cook’s distance statistics was unaffected by model misspecification. The results of this study suggest that Likelihood distance and generalized Cook’s distance are more accurate under more varied conditions in identifying target cases in confirmatory factor models.

Keywords
confirmatory factor analysis, Mahalanobis distance, influence analysis, case deletion, likelihood distance, generalized Cook’s distance, Chi-square difference

Confirmatory factor analysis has become a popular technique for scale development in quantitative research in the social sciences. Researchers have come to rely on fitting confirmatory factor models to determine the extent to which theory about a quantitative scale matches the pattern of responses observed in practice. However, the results of fitting confirmatory factor models can be unduly influenced by one or a few cases in the data (Bollen, 1989). In other modeling frameworks, such as ordinary least square (OLS) regression, there has been extensive development and widespread use of case diagnostics for identifying influential cases in a sample data set. Such has not been the case with confirmatory factor models, which were traditionally fit to descriptive statistics sans individual case data. Several case diagnostics have been proposed and illustrated with confirmatory factor models, but the accuracy of these statistics in detecting simulated target cases has not been systematically examined or compared. Without rigorous and systematic examination, social science researchers have little basis for selecting and using particular case diagnostics in practice. Comparison of the accuracy of case diagnostics in confirmatory factor analysis is warranted to provide practical guidance regarding which statistics are more accurate in identifying target cases under a variety of conditions.

The purpose of this study is to investigate the accuracy of deletion-style case diagnostics, called deletion statistics, available through the R package influence.SEM (Pastore and Altoe, 2017). Our goal is to determine which deletion statistics perform with relatively better accuracy in identifying

¹Quantitative Methods Program, Southern Illinois University Carbondale, Carbondale, IL, USA
²National Center for Assessment, Educational & Training Evaluation Commission, Riyadh, Saudi Arabia

Corresponding author:
Jennifer Koran, Quantitative Methods Program, Southern Illinois University Carbondale, 625 Wham Drive, Carbondale, IL 62901, USA. Email: jkoran@siu.edu
simulated target cases across a variety of conditions in confirmatory factor models. In addition, we compare the accuracy of the deletion statistics to the accuracy of Mahalanobis distance (MD), which is widely used to identify unusual cases in multivariate data. We present a Monte Carlo simulation study that compared the accuracy of Likelihood distance, generalized Cook’s distance, and Chi-square difference with the accuracy of MD. We hypothesized that the case diagnostics would have different accuracy in detecting target cases across variations in sample size, proportion of target cases, and model specification.

Background

Case diagnostics have as their common goal the identification of unusual cases in the data. Unusual cases may be conceptualized as multivariate outliers (Hawkins, 1980), residual outliers (Hildreth, 2013), influential cases (Cook, 1977), or leverage observations (Yuan and Zhong, 2008). Of all these conceptualizations, influential cases are most important due to their demonstrated effects on model results. Multivariate outliers, residual outliers, and leverage observations may or may not be influential cases (for an overview see Flora et al., 2012). Deletion statistics are the most straightforward approach to quantifying case influence. By comparing model results with and without the presence of the ith case in the sample data, deletion statistics provide information about the global influence of the ith case on the results (Cook, 1977; cf Cadigan, 1995; Cook, 1986). The deletion statistic concept is well developed in the context of multiple regression models (e.g. Belsley et al., 1980; Cook and Weisberg, 1982). However, far fewer studies have approached deletion statistics in the context of confirmatory factor models. Studies that have approached this topic generally introduced and illustrated the use of deletion statistics, but did not systematically evaluate or compare their accuracy.

Few deletion statistics have been proposed for use with confirmatory factor models. One such statistic is an adaptation of Cook’s (1977) likelihood distance

\[ LD_i = 2[L(\hat{U}) - L(\hat{U}_{(i)})] \]  

(1)

where \( i = 1, \ldots, N \) and \( L(\hat{U}) \) and \( L(\hat{U}_{(i)}) \) are the model likelihoods in the original sample and sample with the ith case deleted, respectively. Likelihood distance is a model-based measure of the influence of ith case on model fit (Cook, 1977, 1986; Cook and Weisberg, 1982). Likelihood distance provides information about both the magnitude and direction of change in the likelihood function resulting from removing the ith case from the data. Positive likelihood distance values indicate that removing the ith case worsens the model fit, and negative likelihood distance values indicate that excluding the ith case improves model fit.

Chi-square difference is also a model-based case influence measure of overall model fit. Chi-square difference is derived from likelihood distance using the maximum likelihood function

\[ \Delta X = X^2 - X^2_{(i)} \]  

(2)

where \( X^2 \) and \( X^2_{(i)} \) are model Chi-square statistics from the original sample and the sample with the ith case deleted, respectively (Pek and MacCallum, 2011; Zhao and Lee, 1998). The Chi-square difference statistic quantifies case influence on overall model fit by the change in the Chi-square model statistic after removing the ith case. Positive values of Chi-square difference indicate improvement in the model fit associated with the exclusion of the ith case, and negative values of Chi-square difference indicate the exclusion of the ith case worsens the model fit. This marks a distinction between good influential cases and bad influential cases (Pek and MacCallum, 2011). Variations of this approach have also been proposed for other model fit statistics in structural equation modeling, such as the Comparative Fit Index, Root Mean Square Error of Approximation, Tucker–Lewis Index, Akaike’s information criterion, and Bayesian information criterion (Pek and MacCallum, 2011), as well as for indexes of model comparison (Sterba and Pek, 2012).

Variations on Cook’s distance have also been proposed for use with structural equation models (Pek and MacCallum, 2011; Zhao and Lee, 1998). Rather than examine influence on overall model fit, generalized Cook’s distance examines case influence on a set of \( L \) parameters

\[ gCD_i = (\hat{\theta} - \hat{\theta}_{(i)})' \text{VAR}(\hat{\theta})^{-1}(\hat{\theta} - \hat{\theta}_{(i)}) \]  

(3)

where \( \hat{\theta} \) and \( \hat{\theta}_{(i)} \) are vectors of length \( L \) obtained from the original sample and the sample with the ith case deleted, respectively, and \( \text{VAR}(\hat{\theta}_{(i)}) \) is the estimated asymptotic covariance matrix of the parameter estimates obtained from the sample with the ith case deleted. Generalized Cook’s distance can be calculated for any subset of parameters \( L \) instead of the full set of model parameters \( Q \).

Given the quadratic form of generalized Cook’s distance, its lower bound is zero. Thus, generalized Cook’s distance always takes positive values and provides information about the amount of change in the set of \( L \) parameters but not the direction of change. A small value of generalized Cook’s distance indicates a small change in the set of \( L \) parameters estimates associated with the exclusion of ith case from the sample. A large value of generalized Cook’s distance indicates a large change in the set of \( L \) parameters estimates associated with the exclusion of ith case from the sample. The scaled difference (\( \Delta \theta_j \)), an analogue to the statistic that is called DFBETA in regression modeling

\[ \Delta \theta_j = (\hat{\theta}_j - \hat{\theta}_{(i)}) / [\text{VAR}(\hat{\theta}_{(i)})]^{1/2} \]  

(4)
where $\hat{\theta}$ and $\hat{\theta}_{(i)}$ are the parameter estimates obtained from the original sample and the sample with the $i$th case deleted, respectively, is used to identify the direction of the change (Pek and MacCallum, 2011; Zhao and Lee, 1998).

In the absence of the availability or awareness of these few case deletion statistics that exist for structural equation models, researchers routinely examine multivariate data using $MD$. $MD$ is the squared Euclidean distance between the $i$th case and the remaining cases while accounting for the correlation in the data (Mahalanobis, 1936)

$$MD = [(y_i - \bar{y}C^{-1}(y_i - \bar{y})^T]$$

(5)

$$C = \frac{1}{n-1} (y_c)^T (y_c)$$

(6)

where $y$ is the data matrix containing $n$ observations on $p$ variables, $y_c = y_i - \bar{y}$ is the column-centered data matrix, and $C$ is the variance–covariance matrix (De Maesschalck et al., 2000: 2). $MD$ follows a scaled Beta distribution (Veraverbeke and Kottropoulos, 2008), but is often compared against a central Chi-square distribution with $p$ degrees of freedom, where a corresponding low $p$-value means that the $i$th case is a potential outlier (Kline, 2016: 73). In practice, $p$-values associated with the robust $MD$s are used (Rousseeuw and Leroy, 1987). The $MD$ concept has also been applied to factor scores and errors for detecting residual outliers and leverage observations (Yuan and Zhang, 2012).

$MD$ is a popular technique for detecting multivariate outliers and has broad applicability because it does not require fitting a model to the data. However, the multivariate outliers detected using $MD$ are not necessarily influential cases. A number of case influence statistics proposed for use with structural equation models are highly correlated with $MD$, including COVRATIO (Pek and MacCallum, 2011), $\Delta S_i$ (Pek and MacCallum, 2011), and $IND_{CHI}$ (Reise and Widaman, 1999).

The deletion statistics likelihood distance, Chi-square difference, and generalized Cook’s distance’s proposed for use in structural equation modeling have not been rigorously tested for their accuracy. Zhao and Lee (1998) and Pek and MacCallum (2011) merely illustrated the use of these statistics using existing data sets. Lee and Wang (1996) simulated confirmatory factor model data to demonstrate the use of Cook’s distance. However, Lee and Wang used only a single replication and did not attempt to verify the accuracy of the statistic over a variety of conditions. Generally, the deletion statistics proposed for use in structural equation modeling have been introduced and illustrated in the existing literature, but not compared or evaluated for their accuracy.

The purpose of this study is to compare the accuracy of deletion statistics for their ability to identify target cases that could influence results while fitting a single model in confirmatory factor analysis. Furthermore, $MD$ is investigated for the comparative utility of multivariate outlier screening for identifying target cases in confirmatory factor analysis. These case diagnostics are compared under a variety of systematically manipulated conditions while accounting for sampling variability using Monte Carlo (multiple replication) simulation.

**Methodology**

**Data generation**

This study used similar conditions as in Lee and Wang (1996), one of very few prior studies on case diagnostics in confirmatory factor analysis to use simulate data. Target cases were generated from a $N(0, 2.25I_6)$ distribution, where $I_6$ is a $6 \times 6$ identity matrix. Non-target cases were generated using the common factor model $N(0, \Sigma)$, where $\Sigma = \Lambda \Phi \Lambda^T + \Psi$ is the $6 \times 6$ population covariance matrix, $\Lambda$ is the loading matrix with

$$\Lambda = \begin{bmatrix} 0.8 & 0.8 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 & 0.5 \end{bmatrix}$$

and $\Phi = \begin{bmatrix} 1.0 & 0.6 \\ 0.6 & 1.0 \end{bmatrix}$ is the factor correlation matrix, and $\Psi = \text{diag} \{0.36, 0.36, 0.36, 0.75, 0.75, 0.75\}$ is the $6 \times 6$ diagonal matrix of unique variances.

Because sample size is a factor that may be expected to affect the accuracy of case diagnostics, data were simulated with three different sample sizes, 200, 400, and 600. As the proportion of influential cases increases, some deletion statistics may lose their ability to accurately detect individual cases affecting the results. Thus, four proportions of target cases to the number of non-target cases were applied, 0.10, 0.05, 0.02, and 0.01. Sample size and proportion of target cases were fully crossed for a factorial design with 12 conditions. Data were simulated in R v3.5.1 (R Core Team, 2018).

**Case diagnostics**

Likelihood distance, generalized Cook’s distance, and Chi-square difference statistics, as computed in the R package *influence.SEM* (Pastore and Altoe, 2017), were compared. Because model misspecification can affect the identification of target cases, both a correctly specified model, shown in Figure 1, and a misspecified orthogonal factors model, shown in Figure 2, were fit to the simulated data using the R package *lavaan* (Rosseel, 2012). Each factor variance was fixed to identify the model. All other parameters were freely estimated.

The Likelihood distance, generalized Cook’s distance, and Chi-square difference statistics used the boxplot criterion, the default in *influence.SEM* (Pastore and Altoe, 2017), for determining whether a case should be flagged as influential. The boxplot criterion establishes thresholds as a function of the first ($Q1$) and third ($Q3$) quartiles. For statistics whose range is $(-\infty, \infty)$, influential cases stand outside the interval $[Q1 - (Q3 - Q1) \times 1.5, Q1 + (Q3 - Q1) \times 1.5]$ (Pastore and Altoe, 2017). For statistics whose range is $(0, \infty)$,
Methodological Innovations

influential cases are above the threshold \([Q1 - (Q3 - Q1) \times 1.5]\) (Pastore and Altoe, 2017). In addition to the default, results were also obtained with the more traditional threshold where the value 3 replaces the value 1.5 in the boxplot formulas.

Robust MD using the minimum covariance determinant estimator was also calculated for the simulated data using the R package \texttt{mvoutlier} (Filzmoser and Gschwandtner, 2018; Filzmoser et al., 2005) to provide a popular baseline for comparison. The default of 0.5 proportion of the observations was used for minimum covariance determinant estimations. The procedure of Filzmoser, Garrett, and Reimann (2005) was used to determine outliers. In this procedure, the discrepancy statistic measures the departure of the empirical Chi-square distribution from the theoretical Chi-square distribution, but only in the tails (most extreme 2%) where outliers would be expected (Filzmoser, Garrett, and Reimann, 2005). The cut score used in this study for the discrepancy statistic was \(0.222 / \sqrt{n}\), where \(n\) is the sample size.

To assess the accuracy of the case diagnostics under each simulated condition, the miss rate (MR), the ratio of missed target cases to generated target cases, and false alarm rate (FAR), the ratio of flagged non-target cases to generated non-target cases, were computed for each of the statistics for each replication. Results were averaged over 100 replications in R with confidence intervals computed using the standard error of the mean and the inverse \(t\) distribution. Averages were compared across different statistics and across systematic manipulations of the conditions.

Results

Tables 1 and 2 display the results of the simulation study for sample size 200 with the default and traditional thresholds, respectively. Likewise, Tables 3 and 4 display the results for sample size 400, and Tables 5 and 6 display the results for sample size 600. Generally, the likelihood distance and generalized Cook’s distance statistics had lower MRs than robust MD. However, the Chi-square difference statistic often had a higher MR than robust MD, particularly when the fitted model was misspecified. Figure 3 illustrates this trend. There was no consistent trend in MRs as the proportion of the target cases to non-target cases increased. Likewise, there was no consistent trend in MRs across sample sizes.

Robust MD had a far lower FAR than any of the other statistics studied. The Chi-square difference statistic was the most sensitive to model misspecification. When the model was misspecified and the default cutoff was used, the Chi-square difference statistic generally had a higher FAR than any of the other statistics, except for conditions with both small sample size and low proportion of target cases to non-target cases. When the traditional cutoff was used, the Chi-square difference had FARs that were lower than likelihood distance and generalized Cook’s distance, but still notably higher than for MD. Figure 4 illustrates this trend.

FARs for robust MD were remarkably stable as the proportion of target cases increased. FARs for the likelihood distance, generalized Cook’s distance, and Chi-square difference statistics generally decreased as the proportion of target cases to non-target cases increased. Only the Chi-square difference statistic showed occasional inconsistency in this regard. FARs for robust MD decreased as sample size increased. There was no consistent trend in FARs across sample sizes for the likelihood distance, generalized Cook’s distance, and Chi-square difference statistics.

Discussion

In this study, the accuracy of the deletion statistics likelihood distance, generalized Cook’s distance, and Chi-square difference was compared with the accuracy of the multivariate outlier detection statistic MD. Through systematic Monte Carlo simulation, the accuracy of these statistics was evaluated under a greater variety of conditions than in the previous literature. The results have important implications for the use of these case diagnostics with confirmatory factor models in practice.
The results of this study showed that the likelihood distance and generalized Cook’s distance statistics detected target cases more effectively than the Chi-square difference statistic under the software defaults in the influence.SEM package in R. Likelihood distance and generalized Cook’s distance overlooked fewer target cases, and their accuracy
was not undermined by model misspecification. When traditional cutoffs were used in place of the defaults, the Chi-square difference statistic overlooked fewer target cases than likelihood distance and generalized Cook’s distance. Robust MD was far more conservative in identifying target cases than the deletion statistics in this study. Fewer
target cases were identified using robust $MD$ because fewer cases overall, both target and non-target, were flagged by robust $MD$.

These results extend the literature in a direction that is particularly important for practice. Prior studies on likelihood distance, Chi-square difference, and generalized Cook’s distance

Table 5. Miss rates and false alarm rates based on $1.5 \times \text{IRQ}$ cutoff for seven case detection statistics by proportion target cases for sample size 600.

| Statistic | Proportion = 0.01 | Proportion = 0.02 | Proportion = 0.05 | Proportion = 0.10 |
|-----------|------------------|------------------|------------------|------------------|
|           | Miss rate | False alarm rate | Miss rate | False alarm rate | Miss rate | False alarm rate | Miss rate | False alarm rate |
|           | $95\% \text{ CI}$ | $95\% \text{ CI}$ | $95\% \text{ CI}$ | $95\% \text{ CI}$ | $95\% \text{ CI}$ | $95\% \text{ CI}$ | $95\% \text{ CI}$ | $95\% \text{ CI}$ |
| $MD$ | 0.675 0.650 0.700 | 0.675 0.650 0.700 | 0.675 0.650 0.700 | 0.675 0.650 0.700 | 0.510 0.502 0.517 | 0.510 0.502 0.517 | 0.510 0.502 0.517 | 0.510 0.502 0.517 |
| $LD$, $CS$ | 0.297 0.282 0.311 | 0.297 0.282 0.311 | 0.297 0.282 0.311 | 0.297 0.282 0.311 | 0.322 0.316 0.328 | 0.322 0.316 0.328 | 0.322 0.316 0.328 | 0.322 0.316 0.328 |
| $gCD$, $CS$ | 0.300 0.285 0.315 | 0.300 0.285 0.315 | 0.300 0.285 0.315 | 0.300 0.285 0.315 | 0.321 0.315 0.328 | 0.321 0.315 0.328 | 0.321 0.315 0.328 | 0.321 0.315 0.328 |
| $\Delta X^2$, $CS$ | 0.523 0.505 0.541 | 0.523 0.505 0.541 | 0.523 0.505 0.541 | 0.523 0.505 0.541 | 0.536 0.528 0.544 | 0.536 0.528 0.544 | 0.536 0.528 0.544 | 0.536 0.528 0.544 |
| $LD$, $MS$ | 0.283 0.269 0.297 | 0.283 0.269 0.297 | 0.283 0.269 0.297 | 0.283 0.269 0.297 | 0.294 0.288 0.300 | 0.294 0.288 0.300 | 0.294 0.288 0.300 | 0.294 0.288 0.300 |
| $gCD$, $MS$ | 0.280 0.266 0.294 | 0.280 0.266 0.294 | 0.280 0.266 0.294 | 0.280 0.266 0.294 | 0.294 0.288 0.301 | 0.294 0.288 0.301 | 0.294 0.288 0.301 | 0.294 0.288 0.301 |
| $\Delta X^2$, $MS$ | 0.753 0.737 0.769 | 0.753 0.737 0.769 | 0.753 0.737 0.769 | 0.753 0.737 0.769 | 0.754 0.747 0.761 | 0.754 0.747 0.761 | 0.754 0.747 0.761 | 0.754 0.747 0.761 |

Table 6. Miss rates and false alarm rates based on the cutoff $3.0 \times \text{IRQ}$ for seven case detection statistics by proportion target cases for sample size 600.

| Statistic | Proportion = 0.01 | Proportion = 0.02 | Proportion = 0.05 | Proportion = 0.10 |
|-----------|------------------|------------------|------------------|------------------|
|           | Miss rate | False alarm rate | Miss rate | False alarm rate | Miss rate | False alarm rate | Miss rate | False alarm rate |
|           | $95\% \text{ CI}$ | $95\% \text{ CI}$ | $95\% \text{ CI}$ | $95\% \text{ CI}$ | $95\% \text{ CI}$ | $95\% \text{ CI}$ | $95\% \text{ CI}$ | $95\% \text{ CI}$ |
| $MD$ | 0.675 0.650 0.700 | 0.675 0.650 0.700 | 0.675 0.650 0.700 | 0.675 0.650 0.700 | 0.510 0.502 0.517 | 0.510 0.502 0.517 | 0.510 0.502 0.517 | 0.510 0.502 0.517 |
| $LD$, $CS$ | 0.297 0.282 0.311 | 0.297 0.282 0.311 | 0.297 0.282 0.311 | 0.297 0.282 0.311 | 0.322 0.316 0.328 | 0.322 0.316 0.328 | 0.322 0.316 0.328 | 0.322 0.316 0.328 |
| $gCD$, $CS$ | 0.300 0.285 0.315 | 0.300 0.285 0.315 | 0.300 0.285 0.315 | 0.300 0.285 0.315 | 0.321 0.315 0.328 | 0.321 0.315 0.328 | 0.321 0.315 0.328 | 0.321 0.315 0.328 |
| $\Delta X^2$, $CS$ | 0.523 0.505 0.541 | 0.523 0.505 0.541 | 0.523 0.505 0.541 | 0.523 0.505 0.541 | 0.536 0.528 0.544 | 0.536 0.528 0.544 | 0.536 0.528 0.544 | 0.536 0.528 0.544 |
| $LD$, $MS$ | 0.283 0.269 0.297 | 0.283 0.269 0.297 | 0.283 0.269 0.297 | 0.283 0.269 0.297 | 0.294 0.288 0.300 | 0.294 0.288 0.300 | 0.294 0.288 0.300 | 0.294 0.288 0.300 |
| $gCD$, $MS$ | 0.280 0.266 0.294 | 0.280 0.266 0.294 | 0.280 0.266 0.294 | 0.280 0.266 0.294 | 0.294 0.288 0.301 | 0.294 0.288 0.301 | 0.294 0.288 0.301 | 0.294 0.288 0.301 |
| $\Delta X^2$, $MS$ | 0.753 0.737 0.769 | 0.753 0.737 0.769 | 0.753 0.737 0.769 | 0.753 0.737 0.769 | 0.754 0.747 0.761 | 0.754 0.747 0.761 | 0.754 0.747 0.761 | 0.754 0.747 0.761 |

CI: confidence interval; CS: correctly specified model; $gCD$: generalized Cook’s distance; $LD$: likelihood distance; $LL$: lower limit; $MD$: Mahalanobis distance; $MS$: misspecified model; $UL$: upper limit; $\Delta X^2$: Chi-square difference.
introduced and illustrated the use of these statistics but did not compare or evaluate them for their accuracy. Most studies have illustrated the use of these statistics with existing data sets (Pek and MacCallum, 2011; Zhao and Lee, 1998). One study, Lee and Wang (1996), used simulated data, but with only a single replication. These studies did not attempt to systematically evaluate or compare the accuracy of these statistics over a variety of conditions. Practitioners had no empirical evidence to support the use of one statistic over another. This study, although limited, produced some systematic evidence to support the conclusion that likelihood distance and generalized Cook’s distance perform more accurately under default conditions than the Chi-square difference statistic. However, the proportion of incorrectly flagged cases can be reduced by overriding the default with the traditional cutoff and using the Chi-square difference statistic.
Practitioners applying case diagnostic statistics should be aware that likelihood distance, generalized Cook’s distance, and Chi-square difference had a higher rate of flagging cases that should not have been detected (FA R) when compared to **MD**. Thus, once influential cases have been identified, it is necessary to investigate their nature before deciding their fate. This study has shown that, depending on the conditions, a substantial portion of ordinary cases may be identified using deletion statistics. Routine deletion of identified cases is strongly discouraged.

The results should be considered in the light of the Monte Carlo design used in this study. Target cases were generated using the same technique as in Lee and Wang (1996). Target cases were generated from an alternative model in which all variables were unrelated, consistent with an independent guessing or random response style. This model represents only one of many ways that influential cases may arise in the data. Furthermore, the alternative model used to generate the target cases had variances 2.25 times larger than the variances of the cases generated to fit the intended model. The greater variability may have mimicked a mechanism of transposition and other data entry errors that would produce influential cases from a distribution of unusually large variability due to, say, a value of 49 being recorded as 94 or a value of 1.1 being recorded as 0.11, for example. Thus, the alternative model generated target cases that were consistent with certain types of random errors, rather than producing target cases consistent with a systematic alternative arrangement of non-zero relations among the variables.

The fact that **MD** was more conservative in identifying cases than the deletion statistics in this study may be due to the choice of cut score for the statistic in this study. Other cut values for **MD** might produce results closer to those of the other statistics in this study. Nonetheless, the criterion used in this study was the default in the mcoutlier package (Filzmoser and Gschwandtner, 2018). Thus, this criterion is the one most likely to be used in practice.

Although the accuracy of the likelihood distance, generalized Cook’s distance, and Chi-square difference statistics was evaluated under a greater variety of conditions than in the previous literature, both the number of factors and the scope of conditions in a single study must be limited. Future research should consider additional models, models of different sizes, different types of model misspecification, and different methods for generating target cases. The method for generating target cases in this study resulted in some target cases that were indistinguishable from non-target cases due to some overlap in the generating models. This method was similar to that employed in Lee and Wang (1996), a prior study that used simulated data to investigate a deletion statistic in confirmatory factor analysis. Adoption of this approach in this study maintained continuity with the prior literature. The development of an alternative method for generating target cases, while outside the scope of this study, would be a useful avenue for further comparison of deletion statistics in confirmatory factor models.

A number of techniques were outside the scope of this study. Techniques for identifying residual outliers and leverage observations (e.g. Bollen and Arminger, 1991; Yuan and Zhang, 2008, 2012) and for testing local influence (e.g. Cook, 1986; Lee and Wang, 1996) require further study as well. Some case influence statistics avoid the computational intensity of the case deletion mechanism by approximating individual contributions to the model Chi-square statistic (Pregibon, 1981; Reise and Widaman, 1999; Tanaka et al., 1991). Further investigation of the accuracy of these statistics, as well as a comparison of their accuracy to that of deletion statistics, is warranted in further research. Finally, future research should also compare these approaches with forward search algorithms, which may be of interest for identifying groups of influential cases (Mavridis and Moustaki, 2008).

In this study, three case deletion statistics and one multivariate outlier detection statistic were systematically investigated using Monte Carlo simulation in the context of confirmatory factor analysis. The significance of this study is threefold. First, it helped to promote a little known area of research in confirmatory factor analysis. While case diagnostics are routinely used in other types of modeling, such as multiple regression, rarely does the potential influence of unusual cases receive attention in analyses applying confirmatory factor analysis. Second, the study extended the existing literature in which case deletion statistics have been introduced and illustrated but not tested or compared under varied conditions. Third, the results provided practical guidance to practitioners by making empirically based recommendations regarding the choice of one statistic over another.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**ORCID iD**

Jennifer Koran [https://orcid.org/0000-0003-2841-4662](https://orcid.org/0000-0003-2841-4662)

**References**

Belsley DA, Kuh E and Welsch RE (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons.

Bollen KA (1989) *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

Bollen KA and Arminger G (1991) Observational residuals in factor analysis and structural equation models. *Sociological Methodology* 21: 235–262.

Cadigan NG (1995) Local influence in structural equation models. *Structural Equation Modeling* 2: 13–20.

Cook RD (1977) Detection of influential observations in linear regression. *Technometrics* 19: 15–18.
Cook RD (1986) Assessment of local influence. *Journal of the Royal Statistical Society B* 48(2): 133–169.
Cook RD and Weisberg S (1982) *Residuals and Influence in Regression*. New York: Chapman and Hall.

De Maesschalck R, Jouan-Jimbaud D and Massart DL (2000) The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems* 50(1): 1–18.

Filzmoser P and Gschwandtner M (2018) mvoutlier: Multivariate outlier detection based on robust methods (R Package Version 209). Available at: https://CRAN.R-project.org/package=mvoutlier

Filzmoser P, Garrett RG and Reimann C (2005) Multivariate outlier detection in exploration geochemistry. *Computers & Geosciences* 31: 579–587.

Flora DB, LaBrish C and Chalmers RP (2012) Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Frontiers in Psychology* 3: 55.

Hawkins DM (1980) *Identification of Outliers*. London: Chapman and Hall.

Hildreth L (2013) *Residual analysis for structural equation modeling*. PhD Thesis, Iowa State University, Ames, IA. Available at: http://lib.dr.iastate.edu/etd/13400

Kline RB (2016) *Principles and Practice of Structural Equation Modeling* (4th edn). New York: Guilford Press.

Lee SY and Wang SJ (1996) Sensitivity analysis of structural equation models. *Psychometrika* 61: 93–108.

Mahalanobis PC (1936) On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India* 2: 49–55.

Mavridis D and Moustaki I (2008) Detecting outliers in factor analysis using the forward search algorithm. *Multivariate Behavioral Research* 43: 453–475.

Pastore M and Altoe G (2017) influence.SEM: Case influence in structural equation models (R Package Version 21). Available at: https://CRAN.R-project.org/package=influence.SEM

Pek J and MacCallum RC (2011) Sensitivity analysis in structural equation models: Cases and their influence. *Multivariate Behavioral Research* 46: 202–228.

Pregibon D (1981) Logistic regression diagnostics. *The Annals of Statistics* 9: 705–724.

R Core Team (2018) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/

Reise SP and Widaman KF (1999) Assessing the fit of measurement models at the individual level: A comparison of item response theory and covariance structure approaches. *Psychological Methods* 4(1): 3–21.

Rosseel Y (2012) lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48(2): 1–36. Available at: http://www.jstatsoft.org/v48/i02/

Rousseeuw PJ and Leroy AM (1987) *Robust Regression and Outlier Detection*. New York: John Wiley & Sons.

Sterba SK and Pek J (2012) Individual influence on model selection. *Psychological Methods* 17: 582–599.

Tanaka Y, Watadani S and Moon SH (1991) Influence in covariance structure analysis: With an application to confirmatory factor analysis. *Communications in Statistics: Theory and Method* 20: 3805–3821.

Ververidis D and Kotropoulos C (2008) Gaussian mixture modeling by exploiting the Mahalanobis distance. *IEEE Transactions on Signal Processing* 56: 2797–2811.

Yuan KH and Zhang Z (2012) Structural equation modeling diagnostics using R package semdiag and EQS. *Structural Equation Modeling: A Multidisciplinary Journal* 19(4): 683–702.

Yuan KH and Zhong X (2008) Outliers, leverage observations, and influential cases in factor analysis: Using robust procedures to minimize their effect. *Sociological Methodology* 38(1): 329–368.

Zhao Y and Lee AH (1998) Theory and methods: Influence diagnostics for simultaneous equations models. *Australian & New Zealand Journal of Statistics* 40: 345–358.

Author biographies

Jennifer Koran, PhD, is an Associate Professor of Quantitative Methods at Southern Illinois University Carbondale. Her research interests include structural equation modeling, multilevel modeling, growth modeling, and categorical data analysis.

Fathima Jaffari, PhD, is a Senior Measurement Specialist with the National Center for Assessment, Educational & Training Evaluation Commission. Her research interests include structural equation modeling and educational assessment.