

A MODEL IMPLIED INSTRUMENTAL VARIABLE APPROACH TO EXPLORATORY FACTOR ANALYSIS (MIIV-EFA)

KENNETH A. BOLLEN , KATHLEEN M. GATES AND LAN LUO

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

Spearman (Am J Psychol 15(1):201–293, 1904, <https://doi.org/10.2307/1412107>) marks the birth of factor analysis. Many articles and books have extended his landmark paper in permitting multiple factors and determining the number of factors, developing ideas about simple structure and factor rotation, and distinguishing between confirmatory and exploratory factor analysis (CFA and EFA). We propose a new model implied instrumental variable (MIIV) approach to EFA that allows intercepts for the measurement equations, correlated common factors, correlated errors, standard errors of factor loadings and measurement intercepts, overidentification tests of equations, and a procedure for determining the number of factors. We also permit simpler structures by removing nonsignificant loadings. Simulations of factor analysis models with and without cross-loadings demonstrate the impressive performance of the MIIV-EFA procedure in recovering the correct number of factors and in recovering the primary and secondary loadings. For example, in nearly all replications MIIV-EFA finds the correct number of factors when N is 100 or more. Even the primary and secondary loadings of the most complex models were recovered when the sample sizes were at least 500. We discuss limitations and future research areas. Two appendices describe alternative MIIV-EFA algorithms and the sensitivity of the algorithm to cross-loadings.

Key words: exploratory factor analysis, latent variables, model implied instrumental variables, miiv, number of factors, exploratory analysis.

Spearman (1904) marks the birth of factor analysis. He hypothesized a single, general intelligence (“G”) factor influencing the performance on many different tests along with a “specific factor” that was unique to each test and proposed factor analysis to examine these ideas. His “two-factor” (common and specific factors) perspective characterized numerous early applications of factor analysis. But even with an emphasis on a common factor that permeated all observed variables, there was recognition that “group factors,” secondary to the common factor, were in subsets of similar tests (Burt, 1917; Carey, 1916). The group factors stimulated interest in multiple factors rather than factor analysis with only a single, common factor. When the door to multiple factor analysis was opened, it gave rise to the question of how many more? This in turn led toward exploratory factor analysis (EFA) where the number of factors was not specified in advance but were “discovered” from the data in combination with factor analysis algorithms.

Though the procedures for EFA are diverse, there are several characteristics that are common. First, they assume no correlated errors (or unique factors) of the observed variables (or indicators).¹ That is, once the factors are controlled, there are no other sources of covariance among the observed variables that measure the factors. Second, when two or more factors are extracted, EFA allows all

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11336-024-09949-6>.

Correspondence should be made to Kenneth A. Bollen, Thurstone Psychometric Laboratory, Department of Psychology and Neuroscience, Department of Sociology, University of North Carolina at Chapel Hill, 235 E. Cameron Avenue, Chapel Hill, NC 27599-3270, USA. Email: bollen@unc.edu

¹ We use the terms “measure,” “indicator,” and “observed variable” interchangeably to refer to the variables in the data set that are factor analyzed and that “load” on the resulting factors.

indicators to load on all factors apart from the scaling (or reference) indicators when these set the metric of the factors. Third, EFA requires rules to determine when enough factors are extracted to stop the process (e.g., the number of factors equals the number of eigenvalues >1). EFA seeks to use many fewer factors than the number of observed variables. After all, one of EFA's purposes is data reduction so the goal is to summarize numerous observed variables into relatively few factors. Fourth, after the extraction of initial factors most applications of EFA require factor rotation to maximize the high loadings and to minimize the low loadings. Factor rotation algorithms seek solutions that enhance the interpretation of the substantive meaning of rotated solution so that researchers can interpret the meaning of the factors. Fifth, as part of the rotation, EFA requires the analysts to choose between allowing all factors to correlate or be uncorrelated where the former is an oblique solution, and the latter is an orthogonal solution.

A variety of techniques for factor analysis emerged in empirical analyses. These ranged from Spearman's (Spearman, 1904) early use of correlations, partial correlations, and tetrads to Thurstone's (1947) centroid method, and to Lawley and Maxwell's (Lawley and Maxwell, 1963) maximum likelihood estimator. Excellent texts on EFA (e.g., Gorsuch, 1983; Harman, 1976; Mulaik, 2009) are available, and our purpose is not to provide a tutorial on such techniques. But we note that there are common problems shared by most EFA procedures. These include: (1) ambiguity about the number of factors (e.g., Preacher et al. 2013), (2) the factor rotation problem or underidentification (e.g., Darton 1980), (3) how to find simple structure solutions (e.g., Thurstone 1947, Ch. 14), (4) *ad hoc* rules of when factor loadings should be set to zero (e.g., Nunnally 1978, p. 421), (5) assuming that all errors ("unique factors") are uncorrelated even if there are good reasons to allow correlated errors (Bollen 1989, p. 232), (6) nonconvergence of iterative procedures, (7) computational burdens of iterative procedures when researchers apply factor analysis to "Big Data" (e.g., Urban and Bauer 2021), and (8) limited diagnostics for each individual indicator (measure or observed variable).²

Our purpose is to propose a new approach to exploratory factor analysis that attends to these issues by using a model implied instrumental variable (MIIV) framework. The MIIV method for confirmatory factor analysis (CFA) and latent variable structural equation models (SEMs) for prespecified models is covered in other articles (e.g., Bollen 1996; Bollen et al. 2021). However, while instrumental variable approaches exist, a MIIV approach to EFA does not. We created this method not because a MIIV approach completely solves all problems of EFA that we described in the last paragraph, but the MIIV approach can help, and it provides a fresh angle on these problems. For instance, the MIIV approach for estimation is noniterative so there is no issue of having an EFA that does not converge. The MIIV approach also offers diagnostic statistics that apply to individual observed variables (indicators) and their fit in the model. One valuable aspect of these diagnostics is that they take a predictable form when the model has too few factors and hence can aid us in determining the number of factors. Further, there is no rotation problem because the algorithms only apply to identified models and the rotation problem is a consequence of underidentification. Yet another advantage of the MIIV approach is that it has greater robustness to model misspecification than has been established for other factor analysis methods. For instance, the factor loadings of an indicator on a factor in a CFA that are correctly specified are relatively uninfluenced by mistakenly including the indicators of a different factor, such as would occur by having an insufficient number of factors (e.g., Bollen, 2020). Yet another advantage is that we can start with some prior hypotheses about the data even if that does not lead to a complete model specification. Correlated errors ("unique factors") are an example where we might suspect that the errors of the same indicator correlate over time. With our MIIV approach,

² Asparouhov and Muthén (2009) proposed an exploratory structural equation modeling approach that researchers can apply to factor analysis that permits correlated errors and cross-loadings. We will contrast their ESEM to MIIV-EFA in the conclusion.

we can include these in our starting model even if we do not know the number of factors. Our paper is meant as a demonstration of the potential of this novel approach to EFA.

The next section of the paper is a literature review of the use of instrumental variables in factor analysis and how our MIIV approach differs from past efforts. The factor analysis model and assumptions are covered in the subsequent section. This is followed by a section that describes the MIIV approach and algorithm to EFA. We then cover test examples to see how our algorithms perform for several simulation and empirical illustrations. Our conclusions are next in which we describe what we see as remaining problems with our MIIV approach in the hopes of stimulating fresh research. Appendix 1 to the paper describes alternative MIIV EFA algorithms that we tried prior to the algorithm described in the main text. Appendix 2 is on sensitivity of the MIIV-EFA to cross-loadings.

1. Instrumental Variables in Factor Analysis

Before discussing how instrumental variables are utilized for EFA, we first provide some foundational information. Since estimation with instrumental variables has been covered extensively elsewhere (e.g., Angrist & Krueger, 2001; Bollen, 2012; Bowden & Turkington, 1990; Didelez et al., 2010; Sovey & Green, 2011), we simply provide a broad overview here. Instrumental variables are helpful in situations where, at the population or data-generating level, the error for an equation is correlated with one or more covariates in the same equation. When the error correlates with a covariate, ordinary least squares (OLS) regression results in biased and inconsistent estimator of regression coefficients. Estimators that use instrumental variables can create a consistent estimator of coefficients. But to do so, we need to find such instruments.

Instrumental variables are variables that correlate with the endogenous variables on the right-hand side of the equation but do *not* correlate with the composite error. *Model implied instrumental variables* (MIIVs) are observed variables (indicators) that are already part of the model and that meet the conditions to be instruments, that is, they correlate with the right-hand side endogenous variables and are uncorrelated with the composite errors *according to the specified model*. With these variables identified, analysts can use instrumental variable estimators [e.g., two-stage least squares (2SLS)] to estimate the coefficients in a factor model as well as other latent variable models (Bollen, 1996). Estimation with instrumental variables has benefits, such as being “distribution-free” and having greater robustness to structural misspecification than full information estimators such as maximum likelihood (Bollen et al., 2007, 2018; Bollen, 2020).

The overwhelming majority of instrumental variable applications are in regression or simultaneous equations without latent variables. However, instruments apply to factor analysis as well. We are not the first to propose an instrumental variable methodology in EFA. The idea appears to have originated in *Psychometrika* with Madansky (1964), though even he cites antecedences to his work. Nearly two decades later, Hägglund (1982) built on Madansky (1964) paper, and he discussed three different factor analysis by instrumental (FABIN) variables estimators for EFA that he called FABIN 1, FABIN 2, and FABIN 3, also published in *Psychometrika*. In yet another *Psychometrika* paper, Jennrich (1987) proposed improved algorithms to calculate the FABIN 2 and FABIN 3 estimators and added a generalized least squares (GLS) estimator of the factor covariances and unique factor variances that is conditional on the factor loading estimates. Cudeck (1991) described the importance of choosing the best scaling (or reference) indicator for each factor and discussed a method to do so by Du (1986).

Though all these studies used instrumental variables, they differ in varying degrees. To help to explain how these instrumental variable methods and models differ, we begin with a basic equation for a general EFA model,

$$\mathbf{Z} = \boldsymbol{\alpha} + \boldsymbol{\Lambda} \mathbf{L} + \boldsymbol{\varepsilon} \quad (1)$$

where \mathbf{Z} is the vector of N_Z indicators of the factors where N_Z is the number of indicators in the factor analysis. We will use the terms of indicators, measures, or observed variables to refer to \mathbf{Z} . The next symbol, $\boldsymbol{\alpha}$, is a vector of measurement intercepts or constants that are analogous to constants in regression equations. The Λ is the matrix of factor loadings and \mathbf{L} is the vector of factors or latent variables. The $\boldsymbol{\varepsilon}$ is the vector of unique factors, sometimes called the equation “errors.” All authors assume that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\text{COV}(\mathbf{L}, \boldsymbol{\varepsilon}) = \mathbf{0}$ where $E(\cdot)$ is the expected value and $\text{COV}(\cdot, \cdot)$ is the covariance of the variables within the parentheses. As we explain below, the model in Eq. (1) as specified is more general than earlier instrumental variable approaches to EFA approaches in that past instrumental variable methods impose more restrictive assumptions compared to what we use.

Table 1 summarizes the properties of the model and estimators for each of the instrumental variable approaches to EFA. The studies differ in their assumptions about the covariances of the unique factors, the inclusion of intercepts, and whether variables are deviated from their means. The last column includes the model implied instrumental variable (MIIV) method that we propose as a way of contrasting ours with earlier work. The first column of Table 1 gives the properties of the EFA method as defined by each publication. The citations to the publications label the remaining columns, with our new approach being the last column. The first property asks whether the estimator is noniterative (yes or no).³ All these instrumental variable estimators are noniterative and this is seen as a major advantage in terms of computational speed. As researchers begin to analyze large data sets with factor analysis, this computational speed takes on greater importance. This noniterative feature also avoids nonconvergence which sometimes happens for iterative estimators like maximum likelihood.

A second property is whether the model for EFA permits estimating intercepts ($\boldsymbol{\alpha}$) of the measurement equation. Intercepts are useful in examining whether there are constant differences between indicators (measures) with similar ranges of values for a given value of the latent variable. Large differences in intercepts across indicators with similar ranges can point to problematic indicators. All the earlier studies omitted intercepts from their models, though we will include them in ours. Relatedly, all studies besides ours assume that the means of the indicators are zero [$E(\mathbf{Z}) = \mathbf{0}$] as is true for variables that are deviated from their means, and this is shown in Table 1. Another common property of these methods is that they assume that the factors are correlated (“oblique”). Cudeck (1991) assumes uncorrelated common factors only for convenience in parts of the presentation, but for all others Σ_{LL} permits nonzero off-diagonal elements.

For the covariance matrix of the unique factors, all papers except Toyoda (1997) and our MIIV approach assume that they are uncorrelated ($\Sigma_{\varepsilon\varepsilon}$ diagonal). A complication of not allowing correlated errors is that there are frequent situations where researchers know that such correlations are likely. For instance, longitudinal studies frequently use multiple indicators of the same factor measured with the same indicators at each wave of data. The errors of a measure at one time point are often associated with the errors of the same measure at the next time point. Hence, correlated errors are hypothesized. Or sometimes the questions leading to measures have similar wordings and this leads to correlated errors even after taking account of the common factors underlying the indicators. Taking account of such correlated errors as part of the EFA procedure could lead to better fitting models that are easier to interpret. It is for this reason that we propose an EFA method that allows them.

Once researchers estimate the factor loadings ($\hat{\Lambda}$), it is valuable to have an estimate of their standard errors for significance testing and forming confidence intervals. Madansky (1964) did not include standard errors in his study, but Hägglund (1982) and Toyoda (1997) did. Although Jennrich (1987) and Cudeck (1991) focused on the FABIN estimators from Hägglund (1982),

³ There are other noniterative estimators for EFA that do not use instrumental variables. See Albert (1944a,1944b), Bentler (1982), Ihara and Kano (1986), and Kano (1990) for details.

TABLE 1.
Comparison of instrumental variable approaches to exploratory factor analysis.

Properties	Exploratory factor analysis model $\mathbf{Z} = \boldsymbol{\alpha} + \boldsymbol{\Lambda} \mathbf{L} + \boldsymbol{\epsilon}E(\boldsymbol{\epsilon}) = \mathbf{0}$, $\text{COV}(\mathbf{L}, \boldsymbol{\epsilon}) = \mathbf{0}$					
	Publication					
	Madansky (1964)	Hägglund (1982)	Jennrich (1987)	Cudeck (1991)	Toyoda (1997)	MIIV Approach
Noniterative	Yes	Yes	Yes	Yes	Yes	Yes
Intercept ($\boldsymbol{\alpha}$)	$\boldsymbol{\alpha} = \mathbf{0}$	$\boldsymbol{\alpha} = \mathbf{0}$	$\boldsymbol{\alpha} = \mathbf{0}$	$\boldsymbol{\alpha} = \mathbf{0}$	$\boldsymbol{\alpha} = \mathbf{0}$	$\boldsymbol{\alpha} \neq \mathbf{0}$
$E(\mathbf{Z}) = \mathbf{0}$	Yes	Yes	Yes	Yes	Yes	No
Common Factors	Σ_{LL} Correlated	Σ_{LL} Correlated	Σ_{LL} Correlated	Σ_{LL} Uncorrelated	Σ_{LL} Correlated	Σ_{LL} Correlated
Covariances (Σ_{LL})	factors allowed	factors allowed	factors allowed	factors (for convenience)	factors allowed	factors allowed
Unique Factors	$\Sigma_{\epsilon\epsilon}$ diagonal	$\Sigma_{\epsilon\epsilon}$ diagonal	$\Sigma_{\epsilon\epsilon}$ diagonal	$\Sigma_{\epsilon\epsilon}$ diagonal	$\Sigma_{\epsilon\epsilon}$ Correlated	$\Sigma_{\epsilon\epsilon}$ Correlated
Covariances ($\Sigma_{\epsilon\epsilon}$)					allowed	allowed
Standard Errors of $\hat{\Lambda}$	No	Yes	No	No	Yes	Yes
Overidentification	No	No	No	No	No	Yes
Tests of						
Measurement						
Equations						
Procedures to Test # of Factors	No	No	No	No	No	Yes
Equality Constraints	No	No	No	No	Yes	Yes

they did not include their standard errors as part of the discussion. Our MIIV approach includes standard errors of $\hat{\alpha}$ and $\hat{\Lambda}$. These are valuable in assessing whether measurement intercepts and factor loadings significantly differ from zero or other hypothesized values.

Tests are available that can help identify misspecifications in the model. These tests rely on the presence of a greater number of instrumental variables than needed to estimate coefficients. As some background, the coefficients in the equations for indicators in EFA are estimated using instrumental variables. Sometimes, there is more than the minimum number of instrumental variables than needed to estimate them. That is, some indicator equations are overidentified. For instance, consider an indicator that depends on one latent variable. This implies that the scaling indicator of the latent variable is the only variable other than the composite error that is on the right-hand side of the equation. Suppose that there are three instrumental variables for that equation. This means that the equation is overidentified because there are two more instrumental variables than there are coefficients to estimate. In these cases, a Chi-square test for the equation is available that can help to detect possible misspecifications in a model (Bollen 1996, pp 117–118). This test is also referred to as an overidentification test.

Kirby and Bollen (2009) compared a variety of overidentification tests. Though they proposed these overidentification tests in the context of CFA and latent variable SEM, these same overidentification diagnostics apply to EFA. In fact, these are an important part of our approach to building new models through EFA. None of the other instrumental variable EFA approaches discussed overidentification tests, whereas the use of such tests is central to our MIIV-EFA approach.

Among other things, these overidentification tests are essential to our procedures for detecting the number of factors underlying a set of variables. Indeed, one reason for indicators failing the overidentification tests is that an insufficient number of factors are extracted from the data and these indicators load on additional factors. In the next section, we will discuss how we use these tests to add factors in the algorithm that we propose. None of the other EFA instrumental variable papers proposed a procedure with which to determine the number of factors.

Finally, there are situations where equality constraints on parameters are deemed necessary. Toyoda (1997) proposed how instrumental variable methods could include such constraints. We use the MIIV-2SLS from Bollen (1996) and Bollen (2019) for our EFA estimator. Nestler (2014) provides a way to modify this estimator to include equality constraints. The remaining studies do not discuss equality constraints.

In sum, we propose an instrumental variable estimator for EFA that allows intercepts for the measurement equations, correlated common factors, correlated errors, standard errors of factor loadings and measurement intercepts, overidentification tests of equations, and a procedure for determining the number of factors. We also permit simpler structures by removing nonsignificant cross-loadings in a model. Some earlier methods allow some of these features, though none, but ours allows all of them and some features are unique to our approach.

The next section gives more details on the MIIV approach to EFA.

2. MIIV-EFA Approach

In broad terms, the MIIV-EFA approach has a number of steps: (1) choose a starting model, (2) transform latent variables to observed variables (L2O transformation), (3) select scaling indicator(s), (4) determine whether to add another factor (latent variable), (5) check for potential cross-loading variables, (6) estimate and check the new factor model, (7) repeat (3) to (6) until there are none or only one significant overidentification test of indicator equations or no remaining equations that are overidentified, and (8) assess model modifications against substantive knowledge. Except for step (1) and (8), all these steps are automated so that the analyst does not need

to manually implement them. We describe the steps here to better explain how our algorithm proceeds. Below we give more details on each step.

(1) *Choose a starting model*

The first step is to determine where to have the algorithm start. For a complete EFA with no information provided by the user, the algorithm starts at a default model of one factor with uncorrelated errors.⁴ Analysts have the option of prespecifying specific correlated errors. We allow the options of correlated errors in the beginning because sometimes analysts have hypotheses or prior knowledge to incorporate. For instance, suppose that the wording of two indicators of the same factor is similar enough that the analysts expect that the errors of these two indicators are correlated even after controlling their common dependence on a factor. In this situation, the analyst would begin by including the correlated errors and the algorithm would commence from there.

This example illustrates a situation where the hypotheses about the model are partially formulated, but the rest of the analyses is exploratory. We illustrate this condition later with an empirical example. In many other analyses, the default model of one factor without correlated errors will be the starting point.

(2) *Transform latent variables to observed variables (L2O transformation)*

Once the initial factor model is in place, the MIIV-EFA procedure replaces latent variables (factors) with observed variables. Equation (1), $\mathbf{Z} = \boldsymbol{\alpha} + \boldsymbol{\Lambda}\mathbf{L} + \boldsymbol{\varepsilon}$, accommodates all factor analysis models. Because the right-hand side includes the latent factors \mathbf{L} , we cannot directly regress the indicators on the factors. The procedure we use eliminates the latent factor \mathbf{L} , so that the resulting equation contains no latent variables. To do so, we use the procedure from Bollen's (1996) original paper on the MIIV approach to latent variable SEM. This is the latent to observed (L2O) variables transformation.⁵

The MIIV approach has a scaling (reference or anchor) indicator for each factor such that its intercept is zero, its factor loading is one, and it only loads on the factor it scales. In this step, we choose scaling indicator(s) based on our initial impressions of which would be best. In the next step, we have a search process for finding the best scaling indicator for each factor. Partitioning the full set of indicators into the scaling and nonscaling indicators, we rewrite Eq. (1) as

$$\begin{bmatrix} \mathbf{Z}_s \\ \mathbf{Z}_{ns} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\alpha}_{ns} \end{bmatrix} + \begin{bmatrix} \mathbf{I}_s \\ \boldsymbol{\Lambda}_{ns} \end{bmatrix} \mathbf{L} + \begin{bmatrix} \boldsymbol{\varepsilon}_s \\ \boldsymbol{\varepsilon}_{ns} \end{bmatrix} \quad (2)$$

where \mathbf{Z}_s and \mathbf{Z}_{ns} are the vectors of scaling (s) and nonscaling (ns) indicators, the scaling indicators' intercepts are set to zero, and the nonscaling indicators' intercepts are in $\boldsymbol{\alpha}_{ns}$, the \mathbf{I}_s is an identity matrix that represents the factor loading of ones for the scaling indicators, $\boldsymbol{\Lambda}_{ns}$ is the matrix of factor loadings for the nonscaling indicators, and $\boldsymbol{\varepsilon}_s$ and $\boldsymbol{\varepsilon}_{ns}$ are the unique factors of the scaling and nonscaling indicators, respectively.

We write the equation for the scaling indicators as $\mathbf{Z}_s = \mathbf{L} + \boldsymbol{\varepsilon}_s$ and we solve this for the latent variables to find that $\mathbf{L} = \mathbf{Z}_s - \boldsymbol{\varepsilon}_s$. Substituting in for \mathbf{L} , we write the equation for the nonscaling indicators as

$$\begin{aligned} \mathbf{Z}_{ns} &= \boldsymbol{\alpha}_{ns} + \boldsymbol{\Lambda}_{ns}\mathbf{L} + \boldsymbol{\varepsilon}_{ns} \\ &= \boldsymbol{\alpha}_{ns} + \boldsymbol{\Lambda}_{ns}(\mathbf{Z}_s - \boldsymbol{\varepsilon}_s) + \boldsymbol{\varepsilon}_{ns} \end{aligned}$$

⁴ The factor analysis literature sometimes refers to the error or disturbance as the uniqueness term that is composed of measurement error and specific variance. Other literature simply uses error term. We use the term error even though uniqueness term is more in keeping with traditional factor analysis terminology.

⁵ Because our interest lies in factor analysis, we only consider the measurement model (i.e., $\mathbf{Z} = \boldsymbol{\alpha} + \boldsymbol{\Lambda}\mathbf{L} + \boldsymbol{\varepsilon}$), whereas Bollen (1996) also includes the latent variable ("structural") model.

$$= \boldsymbol{\alpha}_{\text{ns}} + \boldsymbol{\Lambda}_{\text{ns}}\mathbf{Z}_s - \boldsymbol{\Lambda}_{\text{ns}} \boldsymbol{\varepsilon}_s + \boldsymbol{\varepsilon}_{\text{ns}}. \quad (3)$$

The last line of Eq. (3) is the L2O variable transformation from the MIIV approach where the latent variables from the original model are substituted out while maintaining identical intercepts and factor loadings to those in the original model in (1) (Bollen, 1996, 2019). The last line of Eq. (3) eliminates the latent factors and has only observed variables on the left-hand side and observed variables and errors on the right-hand side. This completes the second step.

(3) *Select scaling indicator*

In the default situation where the number of factors or correlated errors are not hypothesized, then researchers begin with a one-factor model. More specifically, they have all indicators load on a single factor. A key part of the algorithm is to identify the best scaling indicator for this factor. For this, we estimate a series of one-factor models where each indicator in turn is the scaling indicator. Because there are N_Z indicators, this results in N_Z one-factor models.

The Z_1 indicator is first, and we begin by estimating the one factor model using it as the scaling indicator. In this single-factor model, Eq. (3) becomes

$$\mathbf{Z}_{\text{ns}} = \boldsymbol{\alpha}_{\text{ns}} + \boldsymbol{\Lambda}_{\text{ns}}Z_1 + \boldsymbol{\Lambda}_{\text{ns}}\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_{\text{ns}} \quad (4)$$

where $\mathbf{Z}_{\text{ns}} = [Z_2 \ Z_3 \ \cdots \ Z_{N_Z}]'$, $\boldsymbol{\alpha}_{\text{ns}} = [\alpha_2 \ \alpha_3 \ \cdots \ \alpha_{N_Z}]'$, $\boldsymbol{\Lambda}_{\text{ns}} = [\Lambda_{21} \ \Lambda_{31} \ \cdots \ \Lambda_{N_Z1}]'$ and $\boldsymbol{\varepsilon}_{\text{ns}} = [\varepsilon_2 \ \varepsilon_3 \ \cdots \ \varepsilon_{N_Z}]'$ and these column vectors having dimensions $(N_Z - 1) \times 1$. Next, we would form a second model where Z_2 is the scaling indicator so that

$$\mathbf{Z}_{\text{ns}} = \boldsymbol{\alpha}_{\text{ns}} + \boldsymbol{\Lambda}_{\text{ns}}Z_2 + \boldsymbol{\Lambda}_{\text{ns}}\boldsymbol{\varepsilon}_2 + \boldsymbol{\varepsilon}_{\text{ns}} \quad (5)$$

where $\mathbf{Z}_{\text{ns}} = [Z_1 \ Z_3 \ \cdots \ Z_{N_Z}]'$, $\boldsymbol{\alpha}_{\text{ns}} = [\alpha_1 \ \alpha_3 \ \cdots \ \alpha_{N_Z}]'$, $\boldsymbol{\Lambda}_{\text{ns}} = [\Lambda_{11} \ \Lambda_{31} \ \cdots \ \Lambda_{N_Z1}]'$ and $\boldsymbol{\varepsilon}_{\text{ns}} = [\varepsilon_1 \ \varepsilon_3 \ \cdots \ \varepsilon_{N_Z}]'$. And this would continue rotating through all N_Z indicators in the model.

The MIIVefa R package that we introduce allows the user multiple methods for selecting the scaling indicator. The default method chooses as the scaling indicator that variable that has the fewest number of statistically significant Sargan (1958) overidentification tests for all other indicators combined with the fewest number of nonsignificant factor loadings. In the factor analyses setting, significant Sargan tests indicate that variables may belong to or load on a different factor or that the model has omitted correlated errors with the errors of other indicators (more on this below). Nonsignificant factor loadings suggest it does not have strong relations with other indicators of the same factor, making it a poor scaling indicator. We refer to indicators with either significant Sargan test statistics or nonsignificant factor loadings as problematic variables. If two or more potential scaling indicators are tied on the least number of problematic variables, we choose the one with the highest R^2 when regressed on all other indicators.

The highest R^2 is straightforward in meaning and calculation. The overidentification test for each equation requires more discussion. Kirby and Bollen (2009) compared a variety of equation overidentification tests and found that the Sargan (1958) test had the best performance across different sample sizes, and we use it. The Sargan test statistic (T_{Sargan}) asymptotically follows a chi square distribution with degrees of freedom (df) equal to $(N_Z - 3)$ in the one-factor model with uncorrelated errors. For $N_Z > 3$, the equation is overidentified and testable.⁶ The null hypothesis of the Sargan test is that all MIIVs satisfy the overidentification constraints. If the model is correctly specified, then this is true. If at least one of the MIIVs is uncorrelated with the composite error of the equation, then another interpretation of T_{Sargan} is that it tests whether *all*

⁶ Another way of stating this is that we have more MIIVs than the number of scaling indicators on the right-hand side of the equation.

MIIVs are uncorrelated with the equation composite error versus at least one of the MIIVs are correlated with the regression error. If so, it violates the condition that all MIIVs are uncorrelated with the equation error and suggests model misspecifications.

Matrix expressions for T_{Sargan} are available in several sources (e.g., Kirby, 2009; Bollen, 2021). We explain a simple popular way to calculate T_{Sargan} using the first nonscaling indicator, Z_2 . Below is the Z_2 equation with the estimates of its intercept, factor loading, and error in the one factor model,

$$Z_2 = \hat{\alpha}_2 + \hat{\Lambda}_{21}Z_1 + \hat{\varepsilon}_2. \quad (6)$$

The estimated residual is $\hat{\varepsilon}_2 = Z_2 - \hat{\alpha}_2 - \hat{\Lambda}_{21}Z_1$. Using OLS, regress $\hat{\varepsilon}_2$ on the MIIVs for this equation. Take the $R_{\hat{\varepsilon}_2}^2$ from this equation and multiply it by the sample size, and this is an estimate of T_{Sargan} ($= NR_{\hat{\varepsilon}_2}^2$) that asymptotically follows a Chi-square distribution with $df = (N_Z - 3)$ in the one-factor model. Each overidentified equation has the Sargan test statistic calculated in the same way.

In brief, choose as the scaling indicator, the variable that results in the fewest number of problematic variables. If some indicators have the same number, then choose the variable with the highest R^2 as the scaling indicator.

(4) *Determine whether another factor should be added to the model.*

Count the number of significant overidentification tests and non-significant factor loadings (i.e., problematic variables). If two or more problematic variables exist, then create an additional factor that has these variables loading on it and go to step (3) to select the scaling indicator of that factor. If none or only one indicator has a statistically significant overidentification test, then the next step depends on whether there are one or more factors. If there is only one factor and one or no problematic variable, then stop and report this model as the final one. If there are two or more problematic variables, then create a new factor and load these problematic variables *exclusively* on this new factor. If the model still has more than one problematic variable after the creation of a new latent variable (and selection of scaling indicator in step 3), then go to step (5).

(5) *Check for potential cross-loading variables*

When there is still one or more problematic variables after a new factor is created, cross-load all problematic variables (PV1) on all existing factors and check for *new* problematic variables (PV2) after cross-loading. Remove all unnecessary cross-loadings except those of PV2. If PV2 is equal or less than one, then no need to create a new factor. If PV2 is more than 1, create a new factor and load PV2 *exclusively* on this new factor and move on to step (6).

(6) *Estimate and check the new factor model.*

Re-estimate the model and overidentification tests.

(7) *Repeat (3) to (6).*

Continue adding factors and removing nonsignificant factor loadings until either there is no or only one problematic variable or there are no more overidentified equations.

(8) *Assess model modifications against substantive theory.*

As with most data-driven methods, the results should be treated with caution. Researchers need to assess the modifications such as new factors, cross-loadings, or correlated errors against the substantive knowledge in the field of application. There is no purely mechanical way to do this. Rather, researchers need to assess whether the created factors, the indicators that load on them, or the added cross-loadings are consistent with our understandings of the field. If not, researchers can ignore the implausible modifications in place of alternatives that are more defensible.

Comments on MIIV-EFA Approach

We would like to emphasize several characteristics of the MIIV-EFA approach. As the above procedure describes, we use instrumental variable estimators (Bollen, 1996) to estimate the measurement intercepts and factor loadings. As we explained above, the instruments are variables that correlate with the right-hand side endogenous variables but do *not* correlate with the composite

error. The MIIVs are observed variables (indicators) that are already part of the model and that meet the conditions to be instruments, that is, they correlate with the right-hand side variables and are uncorrelated with the composite error according to the specified model. An algorithm finds the MIIVs for each equation in the model (Bollen, 1996; Bollen and Bauer, 2004; Fisher et al., 2021).

Two-stage least squares (2SLS) and generalized method of moment (GMM) are two instrumental variable estimators proposed for this MIIV approach (Bollen, 1996; Bollen et al., 2014). We use MIIV-2SLS because of the availability of single equation overidentification test statistics (Kirby and Bollen, 2009) that we use when assessing different number of factors, correlated errors, or cross-loadings, though MIIV-GMM is another estimation option. These overidentification test statistics are helpful in finding misspecifications in the model because if the model were correct, then all MIIVs would meet the conditions of instruments. The Sargan (1958) overidentification test statistic is a diagnostic that helps to detect when we have too few factors, omitted cross loadings, or omitted correlated unique factors.

When we have hypotheses such as the number of factors or specific correlated errors, our selection of MIIVs is modified according to these hypotheses. In all cases, we use the algorithm programmed in MIIVsem R package (Fisher et al., 2021) for selection of MIIVs.

3. Illustrations

In this section, we illustrate the MIIV-EFA algorithm. We use both simulation and empirical examples. The simulation examples have the advantage that we know the true data generating model (DGM) while for the empirical example we do not. For those variables that load on more than one factor, we distinguish between the largest or primary loading and the secondary or cross-loading. The performance of the MIIV-EFA algorithm on simulations is evaluated by answering three questions: (1) how many times the true number of factors is recovered, (2) how many times the primary loading structure is recovered correctly, and (3) how many times each individual path is recovered correctly, including both primary and secondary loadings. We begin with eight simulation examples.

3.1. Simulated Data

We simulated data with one, two, three, and four latent factors that contained zero, one, and two cross-loadings for a total of eight simulations. The number of ‘observed’ variables in the simulations ranged from 4 to 20 (see Table 2). Each simulation example was simulated with a sample size of 60, 100, 500, and 1000, which resulted in a total of 32 individual simulated datasets. Each simulation has 500 replications. All latent factors are $N(0,1)$ with a covariance of .5, and the errors are $N(0,1)$. All scaling indicators for each factor in all simulations have factor loadings of 1 while the remaining primary loadings range from .6 to .8, and secondary or cross-loadings range from .3 to .4. With these parameters, the scaling indicators have reliabilities of about .6. We use the R package `mnormt` for multivariate normal distribution generation and provide the code in the online supplementary material. Table 2 summarizes the detailed conditions for each simulation regarding the number of latent factors, the number of variables, and the presence and loadings of cross-loading variables.

Simulations with the same number of factors have all identical parameters except for cross-loading parameters. For example, simulation 7 only differs from simulation 6 in that it adds a cross-loading to Z_{15} on factor 1. We use MIIvefa R package to run the algorithm on all eight simulation conditions (32 individual simulated datasets). In addition, we use a significance level of .05, .01, and .005 for testing the Sargan’s test and loading coefficients when running MIIvefa with

TABLE 2.
Simulation conditions.

	Number of latent factors	Number of total variables	Cross-loading (“secondary loading”) variables	Parameter value of cross-loading (“secondary loading”)
Simulation 1	1	4	NA	NA
Simulation 2	2	8	NA	NA
Simulation 3	2	8	Z ₈	0.4
Simulation 4	3	15	NA	NA
Simulation 5	3	15	Z ₁₅	0.4
Simulation 6	4	20	NA	NA
Simulation 7	4	20	Z ₁₅	0.4
Simulation 8	4	20	Z ₁₅ , Z ₂₀	0.4, 0.3

the purpose to evaluate the potential influences of different significance levels on the performance of the algorithm. This resulted in 48,000 (8 simulation conditions * 500 repetitions * 4 sample sizes * 3 significance levels) MIIvefa outputs for us to evaluate. We start with examining the recovery of the correct dimensionality first.

The number of factors is counted as correct if it matches the number of factors in the true data generating model (DGM). The results of dimensionality recovery by simulation condition, sample size, and significance level are represented in Fig. 1. The x -axes represent sample sizes, and the y -axes represent the number of times the correct number of factors were recovered. Because each simulation condition by sample size has 500 repetitions, a perfect (100%) recovery would be 500, as shown as the maximal y -axis value in Fig. 1.

Perhaps most striking in Fig. 1 is the excellent recovery of the number of factors in the sample sizes of 100 or more. In nearly all replications the correct number of factors results regardless of the significance levels. It is only in sample size of 60 that differences are noticeable. Here using the 0.05 significance level has the best performance compared to the 0.01 and 0.005 levels of significance, and indeed, the differences are notable for some of the models. To the degree that these results generalize, the significance level has only a minor impact when the sample is 100 or more, but for smaller samples the use of the 0.05 significance level would be best in recovery the true number of factors. It is interesting to note that some authors recommend a sample of at least 100 for EFA (Cattell, 2012; Kyriazos, 2018) and our results suggest that at these sample sizes there will be excellent recovery of the number of factors.

As the name primary loading recovery suggests, Fig. 2 shows the number of times that *all* the main or primary loadings of variables in the true DGM are recovered with our algorithm. For example, in the case of simulation 3, if for one repetition all main loadings of Z₁ to Z₇ are recovered but not for Z₈, it would be an incorrect primary loading recovery because of Z₈. Therefore, the y -axis has a maximum number of 500, the number of repetitions for each simulation condition, as the primary loading recovery measure happens at the repetition level.

Compared to Fig. 1, the overall recovery decreased in Fig. 2, especially at smaller sample sizes, which is expected as the primary loading recovery is a stricter measure than recovering the number of factors. Like the patterns present in Fig. 1, the primary loading recovery increased when the sample size increased and when the true DGMs were less complex. In addition, a significance level of 0.05 appears to outperform 0.01 and 0.005 especially at smaller sample sizes. However,

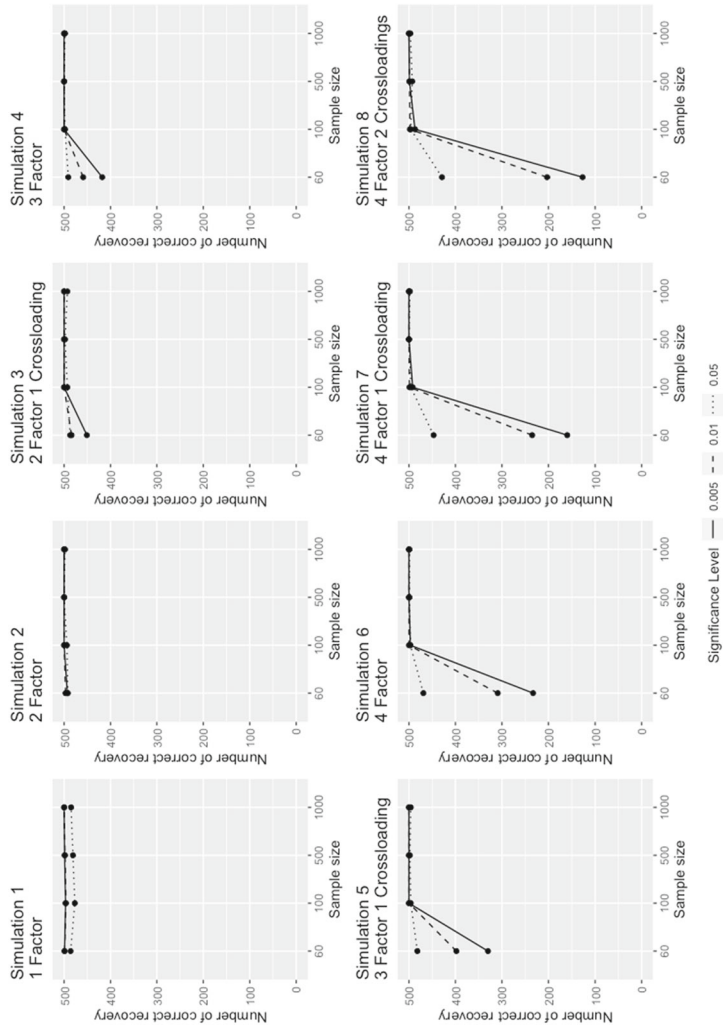


FIGURE 1.
Recovery of correct number of factors using MIYéfa.

this relation was flipped at larger sample sizes for more complex true DGMs as shown in the plots for simulation 7 and 8, where the dotted line ($p=0.05$) became the lowest although all recovery rates are high at the large sample sizes.

The main difference between Figs. 1 and 2 appears to be at $N=100$. While the number of factors recovery became almost perfect after increasing the sample size from 60 to 100, this was not the case for primary loading recovery when the true DGM was more complex (as shown in the plots for simulation 5–8) and when the significance level was at either 0.01 or 0.005. However, for a significance level of 0.05, the primary loading recovery was still almost perfect for the first 6 simulations and was above 95% for the latter more complex DGMs (simulation 7 and 8).

Note that although it is termed primary loading recovery, it is nevertheless a very strict measure—failing to recovery even one main loading in the true DGM would be counted as incorrect. For example, if say, Z_8 was only recovered on the secondary factor but not on the first factor (the primary loading), the recovery would be counted as incorrect. This failing to capture cross-loadings was the main reason for the decrease in recovery rates from Figs. 1 and 2, which we explain in more details in the sensitivity and specificity section.

Figure 3 contains the most stringent standard for evaluating our EFA procedure. It looks at the recovery of both the primary and secondary loadings of the DGM. Note that for true DGMs without cross loadings, the primary loading recovery would be identical to the primary and cross-loading recovery, which means that the plots for simulation 1, 2, 4, and 6 are identical in Figs. 2 and 3.

The general pattern of the influences of sample sizes and significance levels on the primary and cross-loading recovery as shown in Fig. 3 resembles that of Figs. 1 and 2. However, when compared to Fig. 2, the decrease in recovery is more noticeable at $N = 100$. Even at a significance level of 0.05, the exact data structure recovery was at most 70% (simulation 3) and decreased to under 20% when the true DGM became more complex. However, when the sample size was 500 or more, MIIVefa was able to recover the exact data structure almost 100% even for complex DGMs such as simulation 8 that have 20 variables with multiple cross-loadings.

A better recovery at a significance level of 0.01 and 0.005 (around 100%) compared to 0.05 (around 95%) at larger sample sizes also became more noticeable in Fig. 3, which was only evident in Fig. 2 for simulation 8 and is not present in Fig. 1. The three figures show that for different recovery purposes (e.g., number of factors recovery versus primary and secondary loadings recovery) and for different sample sizes, a different significance level could be applied to reach the best recovery. For example, for a large sample size with potentially a complex data structure, MIIVefa at a significance level of 0.01 or 0.005 would be expected to have better primary and cross-loading recovery than a significance level of 0.05; similarly, for a very small sample size, MIIVefa at a significance level of 0.05 would be expected to have a better recovery of the number of factors. In addition, a significance level of 0.05 would be a good default significance level as it had the most consistent results across sample sizes and data structures for all three recovery measures as shown in Figs. 1, 2, and 3.

3.1.1. Sensitivity and Specificity We also calculated sensitivity (“true positives”) and specificity (“true negatives”) for all 8 conditions across all 4 sample sizes. Unlike the previous recovery measures which concern the overall model structure for each repetition, we calculated sensitivity and specificity for each simulation condition by all 500 repetitions. The overall sensitivity results are summarized in Fig. 4. Note that because each simulation condition has different number of *true positive paths* in the true DGMs, the y-axis is now a percentage with a maximum of 100% instead of a raw number of 500 representing the number of repetitions as shown in Figs. 1, 2, and 3. Sensitivity was more influenced by significance levels at smaller sample sizes while a significance level of 0.05 having the highest sensitivity, resembling the pattern in previous figures. Different from Fig. 3 in which the primary and secondary loadings recovery rates are still very low at a

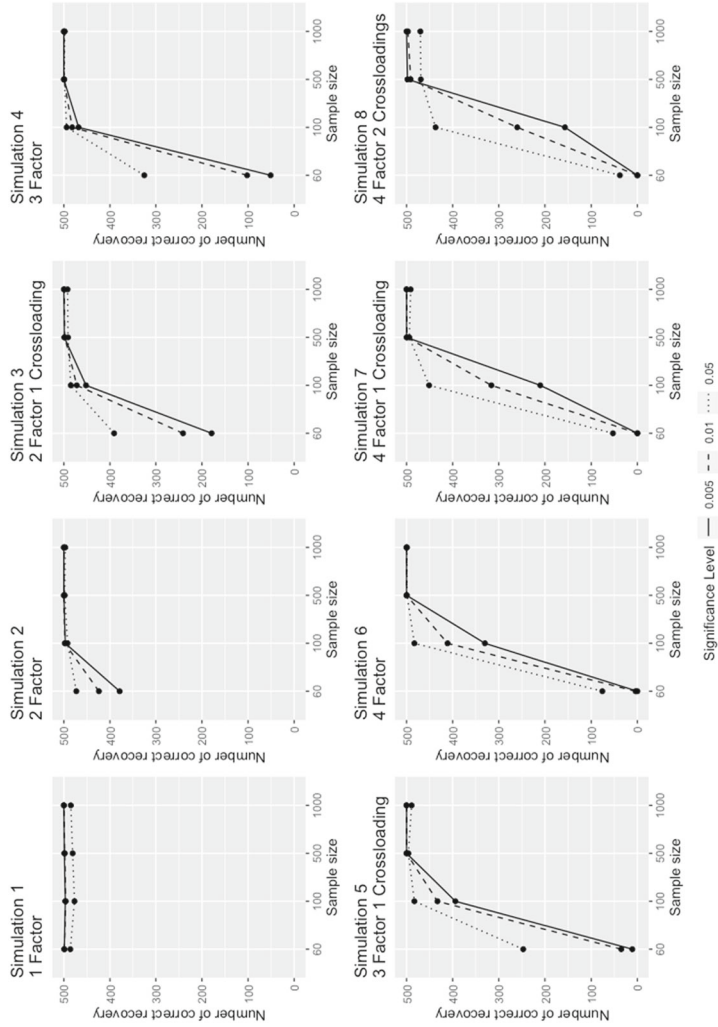


FIGURE 2.
Primary loadings recovery.

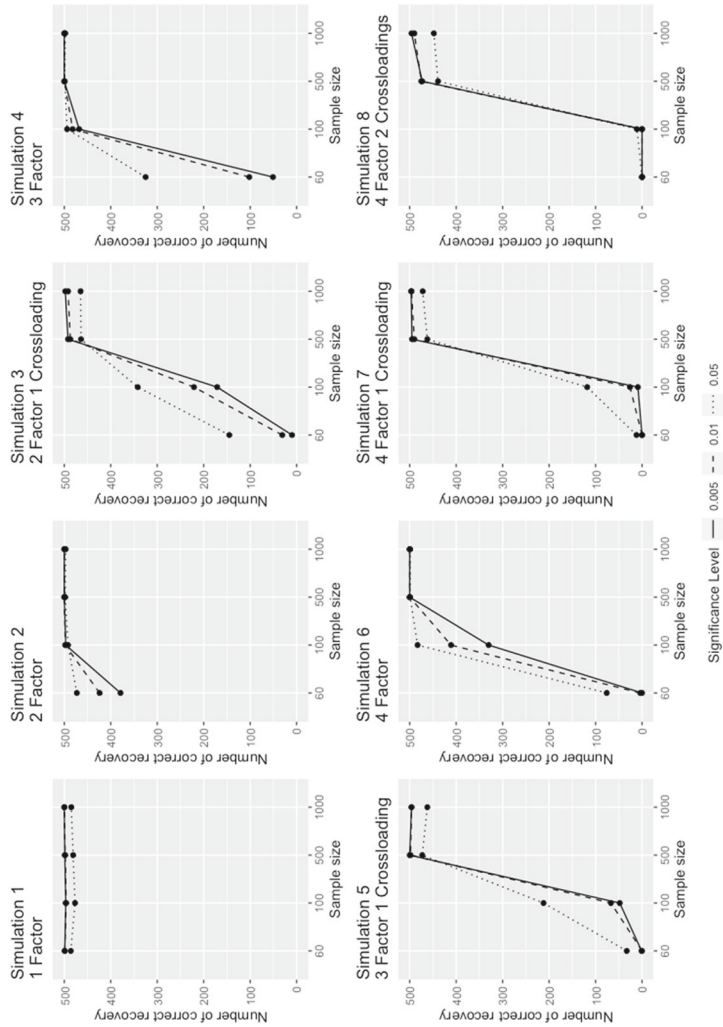


FIGURE 3.
Primary and cross-loadings recovery.

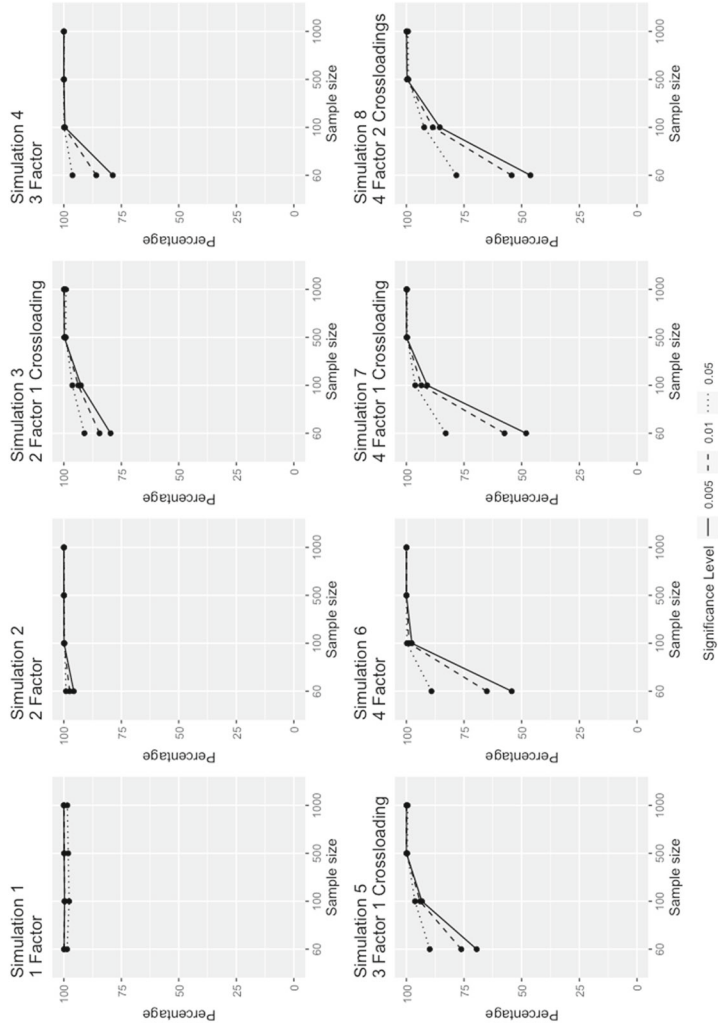


FIGURE 4.
Sensitivity ("true positives") plots for primary and cross-loadings in simulations.

sample size of 100, sensitivity rates are above 85% even for the most complex simulation 8 for all three significance levels. Once the sample size reached 500, MIIVefa had almost perfect 100% sensitivity across all 8 simulations.

Note that the previously discussed primary and cross-loading recovery rates are stricter measures than sensitivity. For example, for one repetition in one simulation, it needs to have a 100% sensitivity, aka all *true positives* in the true DGM are recovered correctly, to have a correct primary and cross-loading recovery. Therefore, sensitivity measures are always higher than or equal to primary and cross-loading recovery rates. The rather large discrepancy between sensitivity measures and primary and cross loading recovery rates, especially at small sample sizes such as 60 and 100 when cross-loadings are present in the true DGMs as shown in Figs. 3 and 4, largely stems from the failure to recover *all* loadings for cross-loading variables. Please refer to Appendix 2 and Fig. 8 for more details on sensitivity for just cross-loading variables.

In addition to sensitivity measures, we also calculated specificity for all eight simulations as summarized in Fig. 5. Note that because simulation 1 is a simple four variable one-factor model which means there is no true *negatives* in the DGM, it cannot be calculated for specificity. Again, like the pattern in previous figures for sensitivity measures and data structure recovery rates, specificity is the lowest and more influenced by significance levels at a sample size of 60. However, once the sample size increased to 100, MIIVefa specificity reached almost 100% for all significance levels. Furthermore, even at a sample size as small as 60, MIIVefa had specificity above 95% across all 8 simulations at a significance level of 0.05. This means that MIIVefa performed very well in terms of not recovering path relations that are not in the true DGMs even for very small sample sizes.

3.1.2. Extra Simulations To explore the performance of MIIVefa under additional conditions, we created two simulation examples based on our simulation 2, which is a simple two-factor, eight-variable condition as introduced earlier. The first modification changes the distribution of observed variables' error terms from normal to log-normal to explore how MIIVefa performs under nonnormality. The first graph in Fig. 6 repeats, for easy comparison, the plot of the recovery of the correct primary loadings under simulation 2 with errors of observed variables from normal distributions as shown in Fig. 1. The second plot in Fig. 6 is the new simulation 2 with errors from a log-normal distribution. In other words, it is the recovery of how many times X_1 to X_4 and X_5 to X_8 were correctly recovered on two separate factors. The most obvious difference is at the smallest sample size of 60 and 100. Under normal errors, the recovery of primary loadings is excellent (more than 80%) across all sample sizes. In contrast, under log-normal distributions the recovery of primary loading also is excellent for sample sizes of 100 or more, but declines greatly for $N=60$. Interestingly, the performance under nonnormality is close to normality even at $N=100$ if the cutoff *p-value* is 0.05, whereas lower *p-values* lead to less accuracy.

Our second modified model is including a pair of correlated errors between X_3 and X_4 to explore the performance of MIIVefa with correlated errors which are mistakenly omitted in the model search procedure. Again, the code to simulate and analyze the secondary simulations is in the online OSF supplementary.

One complication that emerges with correlated errors is what counts as a correct model recovery? It is well known that a CFA model with a single correlated error pair has an equivalent chi-square and degrees of freedom to a model with the exact same structure, except that the correlated error pair is replaced by a new factor that is uncorrelated with the other factors, and has both loadings constrained to one and pointing toward the two indicators that had had the correlated errors. So, if an algorithm chooses a model that replaces the correlated errors with a new factor, it seems inaccurate to count this against the algorithm as a wrong model. After all,

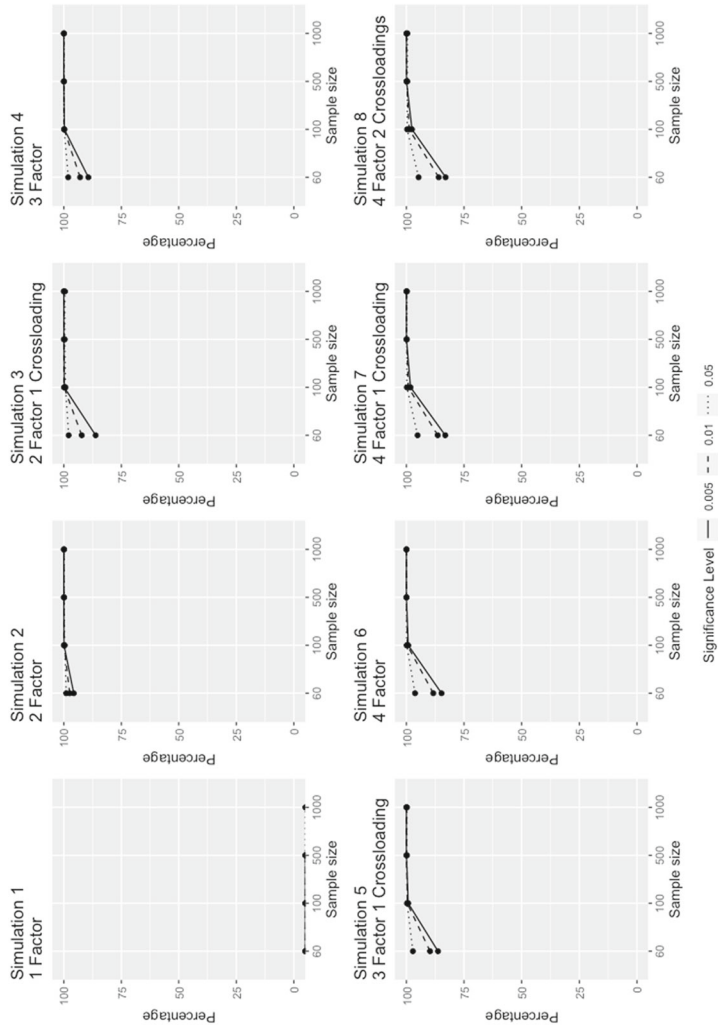


FIGURE 5. Specificity (“true negative”) plots for all simulations.

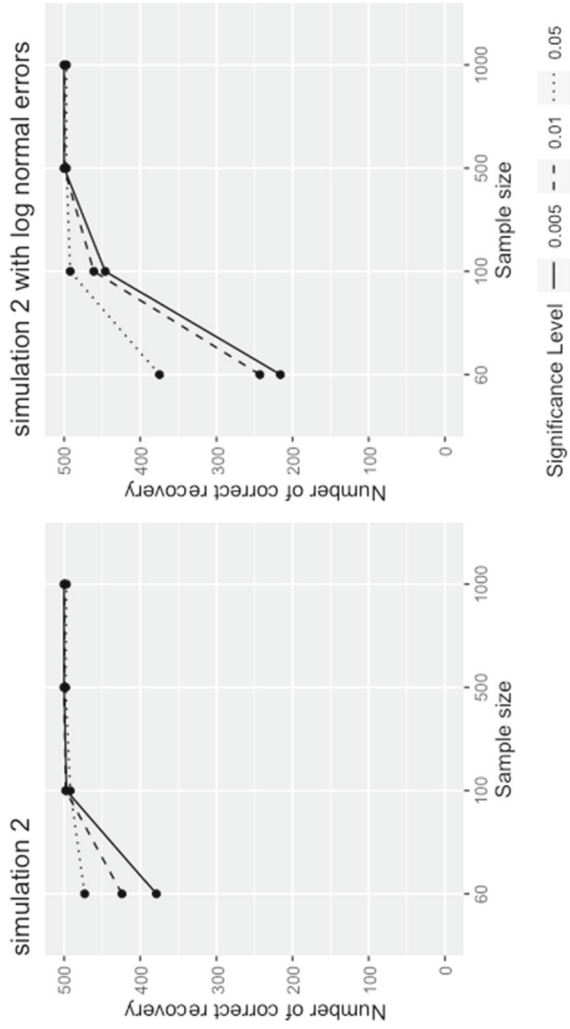


FIGURE 6. Primary loadings recovery with log normal errors.

both models have identical chi square fits and degrees of freedom.⁷ Relatedly, in Bollen (1987) it was found that a one-factor model with correlated error is equivalent to a two-factor model where the indicators with correlated errors load exclusively on a second factor and the other indicators load exclusively on the first factor which suggests another possible model with the same or similar fit.

For these reasons, we separately count the number of times that: (1) a two-factor model is correctly recovered with primary loadings, (2) a three-factor model is correctly recovered with primary loadings, and (3) either a two- or three-factor model is correctly recovered with primary loadings when we count the recovery of the model. Though this complicates the reporting of the results, it appears to be the most complete way to gauge whether the correct number of factors is recovered. Figure 7 reports the results. The leftmost plot shows the number times that the primary loadings of the two-factor model are recovered and this is most common at the smaller sample sizes (60, 100). As the center graph shows, three factors become increasingly more common as the sample size grows. Finally, the two- or three-factor solutions are the most common solutions across all sample sizes with some variations across the choice of p -values. This exploration reveals the complications that correlated errors give rise to in that defining the number of factors because of the presence of equivalent models while at the same time it reveals the ability of the our MIIV EFA algorithm in recovering the equivalent models.

4. Empirical Examples

4.1. Empirical Example #1 Holzinger and Swineford

The first example comes from Holzinger and Swineford (1939), and it consists of mental ability test scores seventh- and eighth-grade children from two different schools. The original dataset contains a total of 26 tests, but a subset of 9 tests is more common in the literature (e.g., Jöreskog, 1969) and the raw data for this example are available as part of the lavaan R package (Rosseel, 2012).⁸ These 9 variables in this empirical data ($N = 301$) are scores for tests on: visual perception, cubes, lozenges, paragraph comprehension, sentence completion, word meaning, speeded addition, speeded counting of dots, and speeded discrimination on straight and curved capitals.

Researchers often use this subset of Holzinger data to test a hypothesized three factor model of mental abilities on visual, textual, and speed tests, without any cross-loading, using CFA. Our algorithm MIIVefa recovered an almost identical model, except that it also recovered one cross-loading: the speeded discrimination of straight and curved capitals on the hypothesized visual factor. The factor loadings are presented in Table 3. We use a significance level of 0.01 for these empirical data ($N = 301$) because our simulation results suggested that a smaller significance level performed slightly better at recovering primary and cross-loadings at larger sample sizes. Note that as previously mentioned, because MIIVefa *cleans up* non-significant factor loadings before printing the final recovered model. The output differs from a traditional EFA output which contains factor loadings of all variables on all factors. The recovered cross-loading is an interesting finding and is empirically reasonable. The speeded discrimination test on straight and curved capitals involves mental abilities of both differentiating between types of capitals visually and being able to do so quickly, and therefore, it naturally loads onto both visual and speed factors.

⁷ Strictly speaking the degrees of freedom match when both factor loadings to the indicators with correlated errors are set to one. In an EFA procedure like ours, we are not introducing such constraints as part of the search procedure.

⁸ The full data set available in *lavaan* also includes demographic data. We only use the nine measures of test scores.

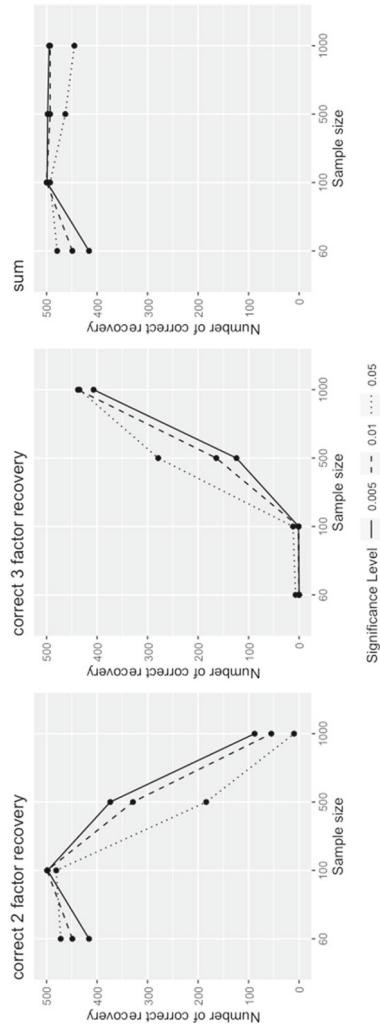


FIGURE 7. Recovery of two-factor, three-factor, or two- or three-factor primary loadings with simulation 2 and correlated errors.

TABLE 3.
MIIVefa recovered model factor loadings for Holzinger data.

	Factor 1 – ‘textual’	Factor 2 – ‘visual’	Factor 1 – ‘speed’
Word meaning	1.000		
Paragraph comprehension	1.079		
Sentence completion	1.197		
Visual perception		1.000	
Cubes		0.632	
Lozenges		0.727	
Speeded counting of dots			1.000
Speeded discrimination straight and curved capitals		0.368	0.665
Speeded addition			0.767

A significance level of 0.01 was employed, and variables with bolded factor loadings were selected as scaling indicators

TABLE 4.
MIIVefa recovered model factor loadings for family planning data.

	Factor 1—(Male perceived accessibility)	Factor 2—(Female perceived accessibility)
Access5	1.000	
Access1		1.132
Access2		1.078
Access3		1.185
Access4	0.904	
Access6	0.963	
Easy1		1.000
Easy2		0.942
Easy3		0.999
Easy4	0.721	
Easy5	0.779	
Easy6	0.776	

A significance level of 0.01 was employed and variables with italicized factor loadings were selected as scaling indicators. Variables labeled the same number had their errors correlated.

4.2. Empirical Example #2 Access to Family Planning Facilities in Tanzania

The second empirical illustration uses a subset of variables from the 1993 Tanzania Accessibility Survey (Mroz et al., 1999). We use measures of subjective assessments of accessibility to family planning clinics in rural communities in Tanzania. Six informants, three females (arbitrarily numbered 1, 2, 3) and three males (arbitrarily numbered 4, 5, 6), were asked to rate both the accessibility (access1 to access6) and how easy it is to get to the clinics (easy1 to easy6). Informant data were collected for 220 rural communities and the rural community is the unit of analysis.

Because the same informant assessed both accessibility and easy to use, it is reasonable to assume that the measurements from the same informant have correlated errors—e.g., access1 and easy1, as they are rated by the same female informant. We ran the algorithm on the data with and without the inclusion of correlated errors between variables rated by the same informant.

When we exclude the correlated errors from the model search, MIIvefa arrived at a seven-factor model—each of the five first factors contained the two assessments by the same informant while for one informant the accessibility and easy were further divided into two separate factors. Table 4 reveals that the MIIvefa solution was simplified by including the correlated errors which resulted in a two-factor model with all male rated assessments on the first factor, and all female rated assessments on the second factor. The results demonstrated that the perceived accessibility of the same clinic differs between male and female informants.

4.3. Empirical Example #3 Fisher Negative Affect Data

The third empirical example comes from Fisher and Boswell (2016). The original data contain 40 individual level intensive repeated measures on 26 mood measures for individuals who had generalized anxiety disorder (GAD) and/or major depressive disorder (MAD) for four times a day for about 30 days. We examined a subset of 8 variables that associated with negative affect: irritable, restless, worried, guilty, anhedonia, hopeless, down, and concentrate, for all 40 individuals (sample size: $mean=113.55$, $min=90$, $max=151$; days: $mean=32.5$, $min=28$, $max=54$). We used a significance level of 0.05 instead of 0.01 as in the previous two empirical examples due to the small sample size, as our simulation results suggested that MIIvefa with a significance level of 0.05 performed better on smaller sample sizes.

Having intensive repeated measures allows researchers to analyze intraindividual differences and to arrive at individual-level models that are person specific. Here we represent three models (individual A, B, and C) recovered by MIIvefa that are the most representative. Factor loadings are summarized in Table 5. Individual A had a clean one-factor model, which means that all eight variables load solely on one negative factor. Individual B had a clean two-factor model, with irritable, restless, worried, and concentrate on one factor and the rest (anhedonia, hopeless, down, and guilty) on the other. The two factors with their corresponding indicators can be interpreted as two separate factors of anxiety and depression. This suggests that while for some people these two factors are internalized as one concept, some other people might see depression and anxiety very differently. Individual C had a two-factor model that is similar to individual B, but with anhedonia being cross-loaded on both factors, which means that although individual C differentiates anxiety and depression similarly as individual B, they have symptoms that could be an indicator for both. This furthermore demonstrates the inter-individual variability that has been found to be present in many psychological processes (Molenaar and Campbell, 2009). For this empirical example specifically, this means that some people are more likely to have comorbidity with GAD and MAD than others (for example, individual A more than individual B and C), and should potentially receive different treatments.

TABLE 5.
MIIVefa recovered model factor loadings for Fisher's negative affect for individuals A, B, and C.

	Individual A	Individual B		Individual C	
	Factor 1	Factor 1	Factor 2	Factor 1	Factor 2
Concentrate	0.419	1.000		1.000	
Irritable	1.191	0.656		1.368	
Restless	1.025	0.529		1.424	
Worried	1.009	0.674		1.104	
Hopeless	1.206		1.071		1.000
Guilty	1.000		1.050		0.803
Down	0.789		1.201		1.260
Anhedonia	1.148		1.000	-0.436	0.378

A significance level of 0.05 was employed and variables with bolded factor loadings were selected as scaling indicators.

5. Conclusions

Analysts have been developing factor analysis methods for well over a century since Spearman (1904) first proposed it. This has resulted in many advances ranging from new procedures for determining the number of factors to novel methods of rotating factor solutions. The possibility of using instrumental variables in exploratory factor analysis has been known since Madansky (1964), but the attention given to instruments in factor analysis has been extremely limited. We proposed an original instrumental variable approach to factor analysis that makes use of MIIVs (Bollen, 1996), overidentification tests (Sargan, 1958), and ways to automate the determination of the number of factors and the presence of cross-loadings. This MIIV-EFA method has many valuable aspects. It uses a MIIV-2SLS estimator which is asymptotically distribution free, so normality or no excess kurtosis of the variables need not be assumed for significance tests of measurement intercepts and factor loadings. It is noniterative so nonconvergence is never an issue. There is no rotation problem because the factor analysis models constructed are identified. In addition, the procedure allows partial input from users who can specify correlated errors when they are theoretically predicted. Furthermore, the MIIV-2SLS has an equation specific overidentification test statistic which provides a local measure of fit that is absent in other approaches to EFA. Finally, we have an R package MIIVefa to implement MIIV-EFA. Beyond presenting the algorithm, we illustrated it with simulations with known DGMs and three empirical examples with unknown DGMs. We found that MIIV-EFA was able to frequently recover the DGMs for the simulation data and came up with plausible models with the empirical examples.

Asparouhov and Muthén (2009) proposed an exploratory structural equation modeling (ESEM) approach that combines features of CFA and SEM with exploratory factor analysis (EFA). Specifically, they permit, an EFA structure to be embedded in a SEM where the EFA part gives rotated factors with cross-loadings and correlated errors that might improve model fit. For a recent overview of ESEM, see van Zyl and ten Klooster (2022). The MIIV-EFA approach differs from ESEM in several ways. First, the MIIV-EFA does not use rotation methods, so there is no need to choose between different methods of rotation. Second, MIIV-EFA estimates single equations rather than the full system of equations. Single-equation estimators like MIIV-2SLS appear more robust to misspecification of models (Bollen et al., 2007), and this should carry over to MIIV-EFA. Relatedly, the MIIV-2SLS estimator is noniterative and does not have the convergence issues that are present with systemwide iterative estimators. In addition, MIIV-EFA has equation specific overidentification tests that can help pinpoint problematic equations. Finally, MIIV-EFA has built in procedures to choose the number of factors.

Though the results for MIIV-EFA are encouraging, we recognize that limitations remain. One limitation is multiple hypothesis testing with the Sargan overidentification test and tests of factor loadings. An overidentification test applies to each overidentified equation and a significance test evaluates factor loadings for each different model that the algorithm creates. If we knew how many tests were to be done in advance, we could apply corrections such as that of Holm (1979) or false discovery rates. But we generally do not know how many tests are implemented until the algorithm ends its search. Our package has the option to control the Type I error rate and we could adjust, but to what value? At the same time, our simulations suggested that higher Type I errors without adjustment often worked better at smaller sample sizes.

Another challenge to EFA in general and to MIIV-EFA as well is the issue of correlated errors. Most EFA methods assumed away correlated errors. MIIV-EFA is less restrictive in that it allows users to specify specific correlated errors that occur with repeated and/or across-group measures or other mechanisms that are hypothesized in advance. However, we do not automate the process of including correlated errors in our algorithm. We experimented with such an option but found that our algorithms were less successful. Part of the challenge is that a pair of correlated errors fits the same as a model that introduces a factor that directly affects the indicators to replace the correlated errors.⁹ Determining whether new factors or correlated errors are appropriate is complicated and requires more study. Our extra simulations illustrated some of the issues.

We also did not discuss how factor score estimates are possible with a model that emerges from MIIV-EFA. The MIIV-EFA procedure provides estimates of the factor loadings and the measurement intercepts. Factor score estimates such as the regression method also need estimates of the covariance matrices of the factors and of the unique factors (“errors”). Bollen (1996, pp 116–117) recommends that the intercepts and factor loadings be entered as fixed values and that the remaining variance and covariance parameters be estimated using maximum likelihood or another system wide estimator that are widely available in SEM software. These would provide consistent estimates of the variances and covariances of the factors and unique factors. Researchers could then use the regression method or other factor score methods that form weighted functions of the observed variables to provide predicted values of all factors in the model. Furthermore, this same approach could give Bollen and Arminger (1991) residuals for individual cases.

Some might wonder how to handle binary or ordinal indicators in factor analysis. Though we did not develop the algorithm for categorical variables, it is a straightforward extension. Bollen and Maydeu-Oliveras (2007), Fisher and Bollen (2020), and Jin et al. (2021) developed a polychoric correlation approach to analyzing mixed types of variables along with asymptotic standard errors and significance tests. Jin and Cao (2018) created an overidentification test analogous to the Sargan (1958) test, but for categorical indicators. These ingredients mean that a MIIV-EFA for categorical and continuous measures should be available.

Finally, we reported the MIIV-EFA algorithm with the best performance in recovering DGMS that we examined. In Appendix 1, we report two other algorithms that we tried but did not work quite as well. We would be surprised if there were no other algorithms that could be devised and that might work even better than what we have proposed. Our goal was to offer a new approach to EFA that uses MIIVs and to demonstrate a “proof of concept.” Our hope is that even the promising algorithm that we developed can be improved upon.

Acknowledgments

We gratefully acknowledge grant support for this research from NIH [1R21MH119572- 01]. Data for this research are available from the cited sources and from the authors.

⁹ The factor loadings for the new factor should be set to 1.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

6. Appendices

6.1. Appendix 1: Alternative Algorithms

The algorithm we report in the text was not the first one we developed. Because it may be instructive to others who seek to develop instrumental variable EFA algorithms, we give a brief description of two other algorithms that we tried before settling on the one in the main text. We also discuss their shortcomings.

Version 1

Our earliest version of the algorithm has several steps that we summarize below:

- (1) Regress each variable on all remaining variables and calculate the R^2 values.
- (2) Start with a one-factor model, using the variable with the highest R^2 value as the scaling indicator.
- (3) Use MIIV-2SLS estimator (the MIIVsem package) to estimate the model.
- (4) Find variables with significant Sargan test statistics. If more than one variable has significant test statistics, create a new factor. If none or one variable is significant, the algorithm stops and prints the current model as the final model.
- (5) Load the variables with significant Sargan test statistics exclusively on the new factor, again use the variable with the highest R^2 as the scaling indicator. Repeat steps 3–5 until no more than one variable has a significant Sargan test statistic.

It is not hard to see that this version is quite simple compared to the current algorithm. But with its simplicity comes limitations. One is that the variable with the highest R^2 might not always be the most appropriate. For example, if the data generating model is a multiple factor model with one variable cross-loaded on all factors, that variable is likely to have a very large R^2 value but questionable as a scaling indicator. Another limitation of this algorithm is that it only considers Sargan test statistics when checking for 'problematic variables' but not factor loadings. This means the algorithm does not remove nonsignificant factor loadings. Finally, this algorithm assumes factor complexity of one so that no cross-loadings are allowed, which is not very realistic for many empirical examples. It is for these reasons that we did not stay with this algorithm.

Version 2

The second version of the algorithm improved version one, and it is very close to the retained algorithm that we reported in the main text. It shares almost all steps, except that it does not include cross loadings. In fact, when no cross-loadings are allowed, this version usually ends up creating extra factors to explain these omitted cross-loading relations.

6.2. Appendix 2: Sensitivity of Cross-Loading Variables

Figure A1 shows the sensitivity for cross-loading variables—the first row is the sensitivity on the primary factor, the second row the secondary loadings on cross-loading factor(s), and the third row the overall sensitivity.

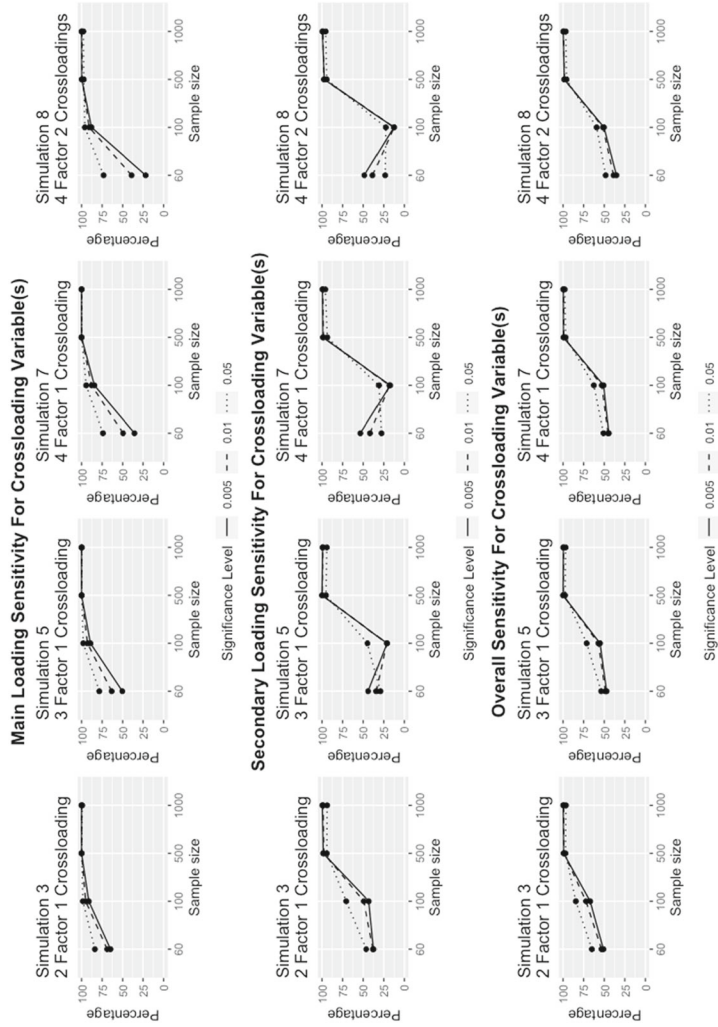


FIGURE 8.
Sensitivity for cross-loading variables.

As represented in Fig. 5, sensitivity for cross-loading variables on their main factors is still high for all three significance levels at a sample size of 100. However, the overall sensitivity dropped at a sample size of 100 across all four simulations because of a lower sensitivity for cross-loadings variables on their corresponding secondary factor(s). This means that it is not that MIIVefa cannot recover *any* cross-loadings at small sample sizes, but that not *all* cross-loadings are always correctly recovered. When dealing with empirical data, the finding for cross-loadings can be aided by substantive background knowledge that simulations do not have. It is worth mentioning that when the sample size is large, MIIVefa works very well at even recovering cross-loadings in addition to primary loadings. Note that the seeming decrease in sensitivity on secondary loadings at $N=100$ is due to the falsely high sensitivity at $N=60$. At $N=60$, as shown in Figs. 2, 3, and 4, the data structure recovery rates and overall sensitivity are both very low. However, when we look at sensitivity for a specific path/relation, it does not account for how accurately the rest of the paths were recovered. For example, at $N=60$ for simulation 5, we could have recovered a model that is very far away from the true DGM and the cross-loading variable is loaded on all factors, which would have resulted in a 100% sensitivity for that specific cross-loading variable. When the sample size increases to 100, again as shown in Figs. 2, 3, and 4, the data structure recovery rates, especially for primary loadings, and sensitivity increase as well, but not too much for secondary loadings on cross-loaded factors. This means that the algorithm is more likely to recover a model that is closer to the true DGM, although it is not uncommon for cross loading variables to be only recovered on the primary factor.

References

- Albert, A. A. (1944a). The matrices of factor analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 30(4), 90–95. <http://www.jstor.org/stable/87882>.
- Albert, A. A. (1944b). The minimum rank of a correlation matrix. *Proceedings of the National Academy of Sciences of the United States of America*, 30, 144–146.
- Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, 15, 69–85.
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16, 397–438. <https://doi.org/10.1080/10705510903008204>
- Bentler, P. M. (1982). Confirmatory factor analysis via noniterative estimation: A fast, inexpensive method. *Journal of Marketing Research*, 19(4), 417–424. <https://doi.org/10.2307/3151715>
- Bollen, K. A. (1987). Structural equation approaches to subjective air quality. In H. S. Koelega (Ed.), *Environmental annoyance: characterization, measurement, and control* (pp. 57–72). Elsevier Science Publishers.
- Bollen, K. A. (1989). *Structural equations with latent variables* (Vol. 210). Wiley.
- Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, 61(1), 109–121.
- Bollen, K. A. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology*, 38, 37–72.
- Bollen, K. A. (2019). Model implied instrumental variables (MIIVs): An alternative orientation to structural equation modeling. *Multivariate Behavioral Research*, 54(1), 31–46.
- Bollen, K. A. (2020). When good loadings go bad: Robustness in factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(4), 515–524.
- Bollen, K. A., & Arminger, G. (1991). Observational residuals in factor analysis and structural equation models. *Sociological Methodology*, 21, 235–62.
- Bollen, K. A., & Bauer, D. J. (2004). Automating the selection of model-implied instrumental variables. *Sociological Methods & Research*, 32(4), 425–452.
- Bollen, K. A., Fisher, Z. F., Giordano, M. L., Lilly, A. G., Luo, L., & Ye, A. (2021). An introduction to model implied instrumental variables using two stage least squares (MIIV-2SLS) in structural equation models (SEMs). *Psychological Methods*, 27(5), 752–772.
- Bollen, K. A., Gates, K. M., & Fisher, Z. (2018). Robustness conditions for MIIV-2SLS when the latent variable or measurement model is structurally misspecified. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(2), 848–859.
- Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., & Chen, F. (2007). Latent variable models under misspecification: Two-stage least squares (2SLS) and maximum likelihood (ML) estimators. *Sociological Methods & Research*, 36(1), 48–86.
- Bollen, K. A., Kolenikov, S., & Bauldry, S. (2014). Model-implied instrumental variable–generalized method of moments (MIIV-GMM) estimators for latent variable models. *Psychometrika*, 79, 20–50.
- Bollen, K. A., & Maydeu-Oliveres, A. (2007). A polychoric instrumental variable (PIV) estimator for structural equation models with categorical variables. *Psychometrika*, 72, 309–326.

- Bowden, R. J., & Turkington, D. A. (1990). *Instrumental variables* (Vol. 8). Cambridge University.
- Burt, C. (1917). The distribution and relations of educational abilities.
- Carey, N. (1916). Factors in the Mental Processes of School Children. III. Factors concerned in the School Subjects. *British Journal of Psychology*, 8, 170.
- Cattell, R. (Ed.). (2012). *The scientific use of factor analysis in behavioral and life sciences*. Springer.
- Cudeck, R. (1991). Noniterative factor analysis estimators, with algorithms for subset and instrumental variable selection. *Journal of Educational Statistics*, 16(1), 35–52.
- Darton, R. A. (1980). Rotation in factor analysis. *Journal of the Royal Statistical Society . Series D (The Statistician)*, 29(3), 167–194.
- Didelez, V., Meng, S., & Sheehan, N. A. (2010). Assumptions of IV methods for observational epidemiology. *Statistical Science*, 25, 22–40.
- Du Toit, A. (1986). The development of a non-iterative method of exploratory factor analysis. Unpublished B.Sc. honors thesis, University of South Africa, Pretoria, South Africa.
- Fisher, A. J., & Boswell, J. F. (2016). Enhancing the personalization of psychotherapy with dynamic assessment and modeling. *Assessment*, 23(4), 496–506.
- Fisher, Z., & Bollen, K. A. (2020). An instrumental variable estimator for mixed indicators: Analytic derivatives and alternative parameterizations. *Psychometrika*, 85(3), 660–683. <https://doi.org/10.1007/s11336-020-09721-6>
- Fisher, Z., Bollen, K., Gates, K., & Rönkkö, M. (2021). MIIVsem: Model implied instrumental variable (MIIV) estimation of structural equation models. R package version 0.5.8, <https://CRAN.R-project.org/package=MIIVsem>.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed., p. 1983). Erlbaum.
- Hägglund, G. (1982). Factor analysis by instrumental variables methods. *Psychometrika*, 47(1), 209–222.
- Harman, H. H. (1976). *Modern Factor Analysis* (3rd ed.). University of Chicago Press.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Holzinger, K. J., & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution. *Supplementary Educational Monographs*, 48, xi + 91.
- Ihara, M., & Kano, Y. (1986). A new estimator of the uniqueness in factor analysis. *Psychometrika*, 51(4), 563–566. <https://doi.org/10.1007/BF02295595>
- Jennrich, R. I. (1987). Tableau algorithms for factor analysis by instrumental variable methods. *Psychometrika*, 52(3), 469–476.
- Jin, S., & Cao, C. (2018). Selecting polychoric instrumental variables in confirmatory factor analysis: An alternative specification test and effects of instrumental variables. *British Journal of Mathematical and Statistical Psychology*, 71, 387–413. <https://doi.org/10.1111/bmsp.12128>
- Jin, S., Yang-Wallentin, F., Bollen, K. A. (2021). A unified model-implied instrumental variable approach for structural equation modeling with mixed variables. *Psychometrika*, 86(2), 564–595. <https://doi.org/10.1007/s11336-021-09771-4>
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(1), 183–202.
- Kano, Y. (1990). Noniterative estimation and the choice of the number of factors in exploratory factor analysis. *Psychometrika*, 55(2), 277–291. <https://doi.org/10.1007/BF02295288>
- Kirby, J. B., & Bollen, K. A. (2009). 10. Using instrumental variable tests to evaluate model specification in latent variable structural equation models. *Sociological Methodology*, 39(1), 327–355.
- Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, 9(08), 2207.
- Lawley, D. N., & Maxwell, A. E. (1963). *Factor analysis as a statistical method*. Butterworths.
- Madansky, A. (1964). Instrumental variables in factor analysis. *Psychometrika*, 29(1), 105–113.
- Molenaar, P. C., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, 18(1), 112–117.
- Mroz, Thomas A., Bollen, Kenneth A., Speizer, Ilene S., & Mancini, Dominic J. (1999). Quality, accessibility, and contraceptive use in rural Tanzania. *Demography*, 36(1), 23–40.
- Mulaik, S. A. (2009). *Foundations of factor analysis* (2nd ed.). Chapman & Hall/CRC.
- Nestler, S. (2014). How the 2 SLS/IV estimator can handle equality constraints in structural equation models: A system-of-equations approach. *British Journal of Mathematical and Statistical Psychology*, 67(1), 353–369.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- Preacher, Kristopher J., Zhang, Guangjian, Kim, Cheongtag, & Mels, Gerhard. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, 48(1), 28–56.
- Rosseel, Yves. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(1), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society*, 26, 393–415.
- Sovey, A. J., & Green, D. P. (2011). Instrumental variables estimation in political science: A reader's guide. *American Journal of Political Science*, 55, 188–200.
- Spearman, C. (1904). 'General intelligence', objectively determined and measured. *The American Journal of Psychology*, 15(1), 201–293. <https://doi.org/10.2307/1412107>
- Thurstone, L. L. (1947). *Multiple factor analysis*. University of Chicago Press.

- Toyoda, H. (1997). A noniterative estimation in confirmatory factor analysis by an instrumental variable method. *Behaviormetrika*, 24, 147–158.
- Urban, C., & Bauer, D. (2021). A deep learning algorithm for high-dimensional exploratory item factor analysis. *Psychometrika*, 86(1), 1–29. <https://doi.org/10.1007/s11336-021-09748-3>
- van Zyl, L. E., & ten Klooster, P. M. (2022). Exploratory structural equation modeling: practical guidelines and tutorial with a convenient online tool for Mplus. *Frontiers in Psychiatry*, 12, 795672. <https://doi.org/10.3389/fpsy.2021.795672>

Manuscript Received: 6 JUN 2023

Accepted: 30 DEC 2023

Published Online Date: 26 MAR 2024