Item factor analysis: Where we've been and where we might be going

Michael C. Edwards

The Ohio State University

June 20, 2008

The 73^{rd} Annual Meeting of the Psychometric Society

Talk Outline

- Some history
- The past decade or two
- The future

Preliminary Matters

Much of this talk is borrowed from Wirth and Edwards (2007).

Preliminary Matters

Much of this talk is borrowed from Wirth and Edwards (2007).

Much of that paper is borrowed from the long line of researchers who have worked on this topic over the years.

A (very brief) History of IFA

The early history of item factor analysis (IFA) followed two parallel courses: one in factor analysis (FA) and one in item response theory (IRT).

Lord and Novick, 1968

Lord and Novick (1968) was one of the earliest places where both frameworks were written down in a formal way. It was also one of the earliest places where relationships between FA and IRT were specified.

Lord and Novick, 1968

Equation 16.8.5 in Lord & Novick is:

$$a_j = \frac{\rho'_j}{\sqrt{1 - \rho'^2_j}} \quad \text{and} \quad b_j = \frac{\gamma_j}{\rho'_j},$$

where *a* and *b* come from an earlier definition of the normal ogive in equation 16.5.2:

$$P_j(\theta) = \Phi[a_j(\theta - b_j)]$$

and γ is a threshold on the underlying response variables and ρ' is the correlation between the underlying response variable and the common factor.

Lord and Novick, 1968

This:

$$a_j = \frac{\rho'_j}{\sqrt{1 - \rho'^2_j}} \quad \text{and} \quad b_j = \frac{\gamma_j}{\rho'_j},$$

Looks an awful lot like this:

$$a_j = \frac{\lambda^*_j}{\sqrt{1 - \lambda^{*2}_j}} \quad \text{and} \quad b_j = \frac{\tau_j}{\lambda^*_j}$$

Lord and Novick, 1968

Early on, there was no way to directly estimate IRT parameters. Instead, parameters from a 1-factor factor analysis of tetrachoric correlations were converted using the equations on the previous slide.

Polychorics

Following the earlier work of Christoffersson (1975) and Muthén (1978), Olsson (1979) introduced a maximum likelihood method for finding the correlations between two or more latent response variables using the proportion of responses in the observed response contingency table.

Polychorics

Although simultaneously estimating all of the thresholds and correlations for a particular set of items is ideal, as the number of items increases there is a corresponding increase in the complexity of the estimation process.

Finding the correlation among latent response variables requires integration. As the number of items increases so does the number of dimensions requiring integration.

Polychorics

Noting the analytic difficulties with simultaneously estimating thresholds and correlations, Olsson (1979) suggested a two-step approach.

In the first step, thresholds are estimated at the univariate level. Then, treating those thresholds as fixed, correlations are estimated bivariately.

Polychorics

Estimating each correlation independently no longer provides “true” maximum likelihood estimates of the correlation matrix and can result in a matrix that is non-positive definite (Song & Lee, 2003) and therefore cannot be inverted.

IRT

- JML
- CML
- MML
- MML/EM

Joint Maximum Likelihood (JML)

Developed by Birnbaum (1968, p. 420), this was the first estimation method capable of directly estimating parameters of IRT models.

This approach used maximum likelihood to estimate all $N + 2K$ parameters simultaneously. There is one “ability" parameter per person (N) and 2 parameters for each of the (K) items (assuming a 2-parameter model). The number of item parameters would change depending on the IRT model(s) being used.

Conditional Maximum Likelihood (CML)

CML takes advantage of the fact that, for the Rasch model, the observed summed score is a sufficient statistic. Using this information, CML estimates the item difficulty parameters *conditional* on the summed scores.

Maximum Marginal Likelihood (MML)

Rather then attempting to estimate all the item and person parameters simultaneously, the MML approach of Bock and Lieberman (1970) *integrates* over the person-specific parameters (this process is known as marginalization) and estimates the item parameters in the marginal distribution.

Maximum Marginal Likelihood (MML)

The rationale to this approach is attributable to Neyman and Scott (1948), who made a distinction between *structural* and *incidental* parameters. In the context of IRT, the item parameters are the structural parameters and the person parameters (θ_i) are the incidental parameters.

MML with an EM algorithm (MML/EM)

Bock and Aitkin (1981) overcame computational difficulties in the MML approach by adapting a strategy now recognized as being equivalent to the EM algorithm.

The Past Decade or Two

IFA has been an extremely active area of research in recent years. The combination of methodological advances and increasingly powerful software have increased applicability and interest.

The Past Decade or Two

IFA has been an extremely active area of research in recent years. The combination of methodological advances and increasingly powerful software have increased applicability and interest.

From a quantitative perspective, this is excellent. From the perspective of someone giving this talk (like me), it is a bit daunting.

SEM-based CFA

A combination of contributions lead to what has been the most popular method for doing IFA in an SEM context. IFA estimation in a CFA framework uses:

- polychoric correlations
- weighted least squares estimation (Browne, 1984)
- post-hoc χ^2 and standard error corrections (Satorra & Bentler, 1994)

SEM-based CFA

Recently there has been a move towards ML estimation for IFA in a number of SEM software packages. This approach is preferable (in my opinion) to the series of "adjustments" that define the list on the previous slide. Unfortunately, depending on which flavor of ML is adopted, new difficulties arise to replace the old.

SEM-based CFA

There have also been recent developments in the area of limited information estimators: The underlying bivariate normality (UBN) approach described in Jöreskog and Moustaki (2001).

It has a similar underlying structure to the polychoric correlation routine, but rather than estimating polychorics it attempts to minimize univariate and bivariate residuals.

SEM-based CFA

SEM-based IFA has a number of strengths including:

- Wide availability of software
- Ease of handling multiple dimensions
- Flexibility of structural model
- Measures of fit
- Mix of continuous and categorical indicators
- Restricted and unrestricted models

EFA

Although it isn't normally done in an SEM framework, it is quite possible to do unrestricted IFA from a factor analytic perspective. Software like CEFA (Browne, Cudeck, Tateneni, & Mels, 2004), among many others, will perform an EFA on polychoric correlations using OLS.

SEM-based CFA

Some of the drawbacks to the SEM approach are:

- Difficult to estimate some IRT models (3PL, nominal)
- New and old estimation challenges
- Estimation issues due to model "complexity"

IRT

The mid to late 1980's saw the addition of another useful IRT estimator, the Marginalized Bayesian (MB) approach of Mislevy (1986)

MB and MML/EM became the "gold-standard" estimators in IRT packages like `BILOG`, `MULTILOG`, and `PARSCALE`.

IRT

As mentioned in Bock and Aitkin (1981) and further developed in Bock, Gibbons, and Muraki (1988), there is nothing that limits the MML/EM approach to unidimensional models.

Unrestricted MIRT models, in addition to the restricted bi-factor model, were implemented in `TESTFACT`.

MIRT

Until fairly recently, `TESTFACT` was one of the only games in town for MIRT modeling. Another program with similar capabilities, but a different framework, was the NOHARM program of Fraser and McDonald (1988).

IRT-based IFA

Advantages:

- Wide variety of item-level models
- Possibility of including priors
- Restricted or unrestricted models
- Retains many common IRT features

IRT-based IFA

Disadvantages:

- Estimation challenges
- Software limitations
- Model fit limited (for now)
- Difficulty with covariates and nesting

Mixed Modeling Framework

De Boeck and Wilson's edited volume *Explanatory Item Response Models* (2004) is an excellent resource for conducting item factor analysis in a mixed modeling framework.

In essence, a generalized or nonlinear (depending on the context) random effects model is estimated where the first level of the model is a item factor analysis model.

Mixed Modeling Framework

De Boeck and Wilson (2004) differentiate *explanatory analysis* and *descriptive measurement*. These two approaches could also be called “modeling” and “scoring”.

The explanatory measurement approach combines both of these objectives without losing the ability to perform either separately. In this sense, the resulting flexibility is very similar to that found in the SEM framework.

Mixed Modeling Framework

The mixed modeling approach to item factor analysis benefits from all the strengths of the mixed model. These include:

- Wide availability of software
- Flexibility in the link functions used
- Mix of continuous and categorical indicators
- Accommodates nesting
- Continuous or categorical predictors and covariates

Mixed Modeling Framework

There are drawbacks to mixed modeling approach to item factor analysis as well:

- Difficult to estimate some IRT models (GRM, 3PL)
- Estimation issues with multidimensional models
- Estimation issues due to model “complexity”
- Confined to confirmatory (or restricted) models

GLLAMM

A related approach that combines features of SEM and mixed models is the Generalized Linear Latent and Mixed Models approach (GLAMM) developed by Rabe-Hesketh and colleagues.

This is a very general framework that, to oversimplify a bit, tends to have the same set of advantages and disadvantages as SEM and mixed models.

The Future?

- Restricted vs. unrestricted
- Compensatory vs. non-compensatory
- One-stage vs. two-stage
- Framework choice
- Estimator choice

The Future?

- Restricted vs. unrestricted
- Compensatory vs. non-compensatory
- One-stage vs. two-stage
- Framework choice
- Estimator choice

The Future?

- Restricted vs. unrestricted
- Compensatory vs. non-compensatory
- One-stage vs. two-stage
- Framework choice
- Estimator choice

The Future?

- Restricted vs. unrestricted
- Compensatory vs. non-compensatory
- One-stage vs. two-stage
- Framework choice
- Estimator choice

The Future?

- Restricted vs. unrestricted
- Compensatory vs. non-compensatory
- One-stage vs. two-stage
- Framework choice
- Estimator choice

The Future?

- Restricted vs. unrestricted
- Compensatory vs. non-compensatory
- One-stage vs. two-stage
- Framework choice
- Estimator choice

The MCMC Wildcard

I have dutifully avoided saying anything about MCMC (and some closely related methods). I've done some work in this area and I have been amazed at how well MCMC can overcome what have been some classic estimation stumbling blocks.

The MCMC Wildcard

I have dutifully avoided saying anything about MCMC (and some closely related methods). I've done some work in this area and I have been amazed at how well MCMC can overcome what have been some classic estimation stumbling blocks.

That said, I don't think maximum likelihood should start making retirement plans any time soon.

The MCMC Wildcard

MCMC is really useful for integration, which is an element in virtually all the frameworks discussed above. In what is probably a further testament to its usefulness, I have seen MCMC used with SEM, IRT, and mixed models (linear and nonlinear).

For the folks in this room (and those like you), I think MCMC will be a viable candidate for estimation.

The MCMC Wildcard

Unfortunately, I do not think MCMC will be something that your "average" social scientist will adopt. We could probably overcome some of the current software issues, but I'm not sure we could completely automate away convergence diagnostics.

I know there are some efforts to do just that (i.e., perfect samplers), but from what I've heard there is considerable work to be done.

The Unexpected

Many of the issues that I've framed as “the future” are either the present or in the immediate future. This leaves the question: What's on the horizon that we don't even know about?

The Unexpected

Many of the issues that I've framed as "the future" are either the present or in the immediate future. This leaves the question: What's on the horizon that we don't even know about?

Time will tell, but I'm excited to see where we stand in 2018.

If I was a betting man...

In 2018 I think we'll see:

If I was a betting man...

In 2018 I think we'll see:

1. All frameworks alive and kicking for IFA

If I was a betting man...

In 2018 I think we'll see:

1. All frameworks alive and kicking for IFA
2. Greater comfort with overlap among the frameworks

If I was a betting man...

In 2018 I think we'll see:

1. All frameworks alive and kicking for IFA
2. Greater comfort with overlap among the frameworks
3. ML with some sort of stochastic integration will be the dominant estimation method

If I was a betting man...

In 2018 I think we'll see:

1. All frameworks alive and kicking for IFA
2. Greater comfort with overlap among the frameworks
3. ML with some sort of stochastic integration will be the dominant estimation method
4. MCMC will still be around, but largely limited to methodologists

If I was a betting man...

In 2018 I think we'll see:

1. All frameworks alive and kicking for IFA
2. Greater comfort with overlap among the frameworks
3. ML with some sort of stochastic integration will be the dominant estimation method
4. MCMC will still be around, but largely limited to methodologists
5. Me giving a talk about how wrong I was back in 2008

The End

Thanks.

edwards.134@osu.edu