Estimating measurement error in longitudinal datasets

An important but tricky problem

Harvey Goldstein, University of Bristol
Michele Haynes, Australian catholic University
George Leckie, University of Bristol

What’s the problem?

- Failure to adjust for Measurement error(ME) in covariates where we have ‘unreliable’ measures can lead to serious parameter biases.
- Several procedures for handling ME have been available:
  - Fuller using moment based estimators
  - SIMEX using successive approximations
  - Bayesian joint models related to missing data modelling
- All these assume knowledge of the distribution of the measurement errors (e.g., normally distributed measurement errors with known variance).
- Little guidance on how to estimate the distributions.
- Often a special problem for longitudinal data where the use of conditional models with errors in predictors is common.
- Will illustrate with educational data

A basic ME model

- Consider the simple regression MOI (we elaborate later)
- $Y = \alpha + \beta X + e,$ $\qquad e \sim N(0, \sigma_e^2)$
- If there is random measurement error *m*:
  - we write the observed covariate $x = X + m,$ $\qquad m \sim N(0, \sigma_m^2)$
  - $cov(m, X) = 0$
  - and possibly the observed response $y = Y + \delta$ $\qquad \delta \sim N(0, \sigma_\delta^2)$
  - Note 'Berkson' model assumes $cov(m, X) = \sigma_m^2$ with appropriate modifications.
- The reliability $R = {}^{var(X)}/_{var(x)}$
  - Estimates of *R* often not available
  - Software not easily available even if *R* known
  - Awareness of the problem limited
- If x used instead of X the regression coefficient β will be downwardly biased in absolute value, by a factor *R*, i.e. we will estimate $R\beta$.

An algorithm to handle ME for known *R*

- Details in Goldstein et al (2014). We write the full model with independent normal residuals, and x_1 are covariates with error.

  $Y = \alpha + \beta X_1 + e,$

  $X_1 = X_2\alpha + \gamma_2$ Note that we need this line

  $x_1 = X_1 + \gamma_1$

- X_2 are measured without error. $X = \{X_1, X_2\}$. *R* is assumed known.
- We use MCMC Metropolis steps to sample parameters.
- Incidentally we can also handle missing values using imputation; ME is a kind of missing value.
- At each step the posterior is computed from the joint likelihood (product) from the second line and the MOI (first line).
- This results in a single chain for MOI that can be used for inference in the usual fashion.
- Note especially that we can directly incorporate interactions involving X_1 variables (not available with moment type estimators) and extend to repeated measures and other multilevel structures.

Estimating *R*

- Assume no replication possible as with educational tests or opinion scales.
- **Internal consistency estimates** (e.g. coefficient alpha)
  - For simplicity consider binary (0,1) responses (p_{ij}) for individual i and item *j*, summed to form a score to be treated as a covariate, $x_i = \sum_{j=1}^{k} p_{ij}$
- Suppose we divide the test items at random into two (approximately) equal groups and assume that for each testee (ie conditional on their characteristics – ability etc.) $\forall j \neq k|i,\ p_{ij} \perp p_{ik}$ – an assumption of conditional or local independence.
- We can treat the scores from each such group as an independent replicate and hence obtain an estimate for the 'half test', between-replicate covariance. Thus for the whole test score, an equivalent estimate of the measurement covariance would simply be four times this value.
- $\alpha = (k/(k-1))(1 - (\sum_{j=1}^{k} P_j(1 - P_j))/\sigma_x^2)$ (1)
- where P_j is the proportion of the sample with correct answers to item *j* and σ_x^2 is observed variance of the sum score

More on α

- Items treated as a 'random' sample from a universe of them
- The conditional independence assumption, however, is crucial. In the formula(1) we have a numerator $P_j(1 - P_j)$ for the variance of true values that essentially assumes independence across items so that, if we add positive covariances (as might be expected), then (1) will tend to overestimate reliability.
- Simulations under an assumed (probit response) model confirm this.

Simulations for α

- Generate data using a a simple item response model
- $\pi_{i,2} = \pi_{i,1}(\int_{-\infty}^{\theta_i - \alpha_2 + c} \phi(t)dt) + (1 - \pi_{i,1})(\int_{-\infty}^{\theta_i - \alpha_2 - c} \phi(t)dt))$
- etc for $j > 2$. For $j = 1,\ c = 0$. 30 item test.

$\boldsymbol{\rho_{12} = \rho_{23} = 0.2}$	C=0	C=0.02
R(true)	0.81	0.59
R(α)	0.82	0.88

Instrumental variable estimators

- Using previous model

  $Y = \alpha + \beta X_1 + e,$ (1)

  $X_1 = X_2\alpha + \gamma_2$

  $x_1 = X_1 + \gamma_1$

- Define an instrumental variable (IV) Z where we assume a linear model relating x, Z ,namely
- $x_i = \gamma_0 + \gamma_1 Z_i + e_z, \ \ \gamma_1 \neq 0,$ $\widehat{x_{1i}} = \gamma_0 + \gamma_1 Z_i$
- Two stage LS substitutes $\widehat{x_{1i}}$ for X_1 in (1) to obtain adjusted (unbiased) estimate of β that is then compared with biased estimate of β to calculate *R*.
- Note that we require $Z \perp e$ which implies $\rho_{ZY} = \rho_{XZ}\rho_{XY}$.
- Note especially that the IV should be highly correlated with X_1.

IV choices

- Grouping estimators
- A commonly advocated, but unsatisfactory, IV method is the so called grouping procedure, Wald (1940), Bartlett (1949) , Durbin (1954).
- Rank and group x_1 into 2 or more then use group membership as IV
- Only works when grouping on basis of *true* scores X_1 or equivalently when measurement errors so small that their intervals around observed do not overlap
- Still advocated in text books, especially economics ones e.g. Johnston, 1972, Cameron and Trivedi, 2005.

IV choices ctd.

- With cross sectional data while we can often get high correlations the assumptions of orthogonality and the necessary correlation structures are difficult to satisfy. So:
- Using distal score as IV in longitudinal data
- Typically not correlated with ME. But needs to satisfy $\rho_{ZY} = \rho_{XZ}\rho_{XY}$
- Simulation for different known *R* and correlations:

$$\begin{pmatrix} 1 & & \\ 0.5 & 1 & \\ q & 0.5 & 1 \end{pmatrix}$$

Here $\rho_{ZY} = \rho_{XZ}\rho_{XY}$ is satisfied when *q* = 0.25 (i.e., autoregressive structure)

Distal IV simulation

N	R_X	$q = \rho_{ZY}$	$50\text{th}(\hat{R}_2)$	$2.5\text{th}(\hat{R}_2)$	$97.5\text{th}(\hat{R}_2)$
100	0.800	0.25	0.768	0.210	3.249
1000	0.800	0.25	0.794	0.630	1.072
10000	0.800	0.25	0.802	0.739	0.868
1000	0.700	0.25	0.709	0.542	0.949
1000	0.900	0.25	0.902	0.701	1.235
1000	0.800	0.20	0.870	0.646	1.342
1000	0.800	0.30	0.583	0.452	0.742

IV example

- Australian school tests (NAPLAN) : years 3,5,7 for 2011 (year 3) cohort.
- Estimating year 5 Maths subtest score reliabilities.

Table 3. Reliability estimates comparing Cronbach's alpha with observed year 5 patterns and IV methods based upon year 3 scores as distal in the regression of year 7 on year 5 scores.

	Algebra, Function & Pattern (4 items)	Measurement, Chance and Data (13 items)	Number (13 items)	Space (10 items)	Numeracy Total (40 items)
Year 5 IV	0.403	0.648	0.625	0.528	0.793
Year 5 coefficient alpha	0.449	0.678	0.681	0.576	0.864

Choosing IV variables

Reliability estimates for measurement test score (13 items) at 5 years using different combinations of IV variables. All models are additive linear regression models.	
IV variables predicting year 5 measure score	Reliability estimate
Year 3 measurement	0.648
Year 9 measurement	0.585
Year 3 total numeracy	0.722
Year 3 total numeracy + year 3 measurement	0.645
Year 3 total numeracy + year 9 measurement	0.616
Year 3 measurement + year 9 measurement	0.655

Some tentative conclusions

- If we use a test score as IV that is measured at the same time as the target variable we find that the estimate of reliability tends to increase.
- In other words it suggests the observed test score contains less measurement error than using a distal test score, so we can infer that we are not fully correcting for ME.
- Grouping estimators should not be used.
- Simulations that allow positive correlations between successive items show that α will be biased upwards.
- Sensitivity analysis should be carried out with a range of reliability values.
- More empirical work needed.

Thank you for listening