Robust Confidence Intervals for Effects Sizes in Multiple Linear Regression

Paul Dudgeon

Melbourne School of Psychological Sciences
The University of Melbourne. Vic. 3010
AUSTRALIA
dudgeon@unimelb.edu.au

Modern Modeling Methods (M^3) Conference
University of Conneticut
20-21 May 2014

Effect Sizes for IVs in Multiple Linear Regression

- Unstandardised regression coefficients (B_j) in linear regression are natural effect sizes.

- Additional effect sizes measures:
  1. Standardized regression coefficient (B_j^*).
  2. Semipartial correlation (sr_j).
  3. Improvement in R-squared (ΔR_j^2).

- These three effect size measures are the primary focus of this talk.
  - Develop confidence intervals robust to assumption violations.
  - Structural equation modelling as a unified framework for deriving and calculating standard errors.

Previous Research

- Limited previous research in this area.
- Algina and colleagues (2001, 2007 2008, 2010) investigated various methods for deriving confidence intervals for the improvement in R-square.
- Aloe & Becker (2012) derived a standard error for the semipartial correlation for use in meta-analysis.
- Yuan & Chan (2011) developed standard errors for standardized regression coefficients.
- With one exception, these approaches all rest on standard regression assumptions holding.

The Linear Regression Model

- The linear regression model for J independent variables measured on $n = 1, \ldots, N$ individuals is given by:

$$Y_n = B_0 + B_1 X_{n1} + B_2 X_{n2} + \ldots + B_J X_{nJ} + e_n,$$

- Estimated values for the regression coefficients can be obtained by the ordinary least squares estimator

$$\min_{(\hat{B}_0, \ldots, \hat{B}_J)} = \sum_{n=1}^{N} (Y_n - \hat{Y}_n)^2$$

where

$$\hat{Y}_n = B_0 + B_1 X_{n1} + B_2 X_{n2} + \ldots + B_J X_{nJ}$$

Deriving Focal Effect Sizes

- All three effect sizes can be related to the unstandardized regression coefficient.

1. Standardized regression coefficient.

$$B_j^* = B_j \times \frac{s_{X_j}}{s_Y}$$

2. Standardized regression coefficient.

$$s\,r_j = B_j^* \times \sqrt{1 - R^2_{(j)}}$$

3. Improvement in R-squared.

$$\Delta R_j^2 = R^2 - R_j^2 = sr_j^2$$

where $1 - R^2_{(j)}$ is the tolerance of the j-th independent variable, and R_j^2 is the reduced model coefficient of determination when X_j is removed.

Linear Regression and SEM

- LS estimator for linear regression is equivalent to the maximum likelihood (ML) estimator.

- Any linear regression analysis can therefore be readily expressed as a (saturated) structural equation model (SEM).

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}_B)$$

where the set of freely estimated model parameters equals

$$\boldsymbol{\theta}_\beta = [\beta_1, \ldots, \beta_J, \sigma_\varepsilon^2, \mathit{vech}(\boldsymbol{\Sigma}_\mathbf{X})]$$

and $\mathit{vech}(\bullet)$ is the matrix operator extracting non-duplicated elements of a symmetric matrix into a vector.

Linear Regression and Correlational Structures

- For linear regression, the SEM specification is straightforward:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}_{LR}) = \left[\begin{array}{c|c} \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\beta} + \sigma_{\varepsilon}^2 & \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\mathbf{X}} \\ \hline \boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\beta} & \boldsymbol{\Sigma}_{\mathbf{X}} \end{array}\right]$$

- Many covariance structures can be equivalently estimated as correlational structures (with appropriate constraints).

$$\boldsymbol{\Sigma}_{Y\mathbf{X}} = \mathbf{D}_{\sigma}\,\boldsymbol{P}(\boldsymbol{\theta}^*_{LR})\,\mathbf{D}_{\sigma},$$

where

$$Diag[\boldsymbol{P}(\boldsymbol{\theta}^*_{LR})] = \boldsymbol{I}_k$$

is the imposed constraint function, $\mathbf{D}_{\sigma}$ is a diagonal scaling matrix, and $\boldsymbol{\theta}^*_{LR}$ contains parameters in a standardized metric.

SEM for Standardized Regression Coefficients

- Standardized regression coefficients can be obtained therefore as:

$$\boldsymbol{\Sigma}_{Y\mathbf{X}} = \mathbf{D}_{\sigma} \left[\begin{array}{c|c} \boldsymbol{\beta}^{*\prime} \boldsymbol{P}_{\mathbf{X}} \boldsymbol{\beta}^{*} + \sigma_{\varepsilon}^{2} & \boldsymbol{\beta}^{*\prime} \boldsymbol{P}_{\mathbf{X}} \\ \hline \boldsymbol{P}_{\mathbf{X}} \boldsymbol{\beta}^{*} & \boldsymbol{P}_{\mathbf{X}} \end{array} \right] \mathbf{D}_{\sigma}$$

with the constraint function

$$\sigma_{\varepsilon}^{2} = 1 - \boldsymbol{\beta}^{*\prime} \boldsymbol{P}_{\mathbf{X}} \boldsymbol{\beta}^{*}$$

- The set of freely estimated model parameters now equals:

$$\boldsymbol{\theta}_{\beta *} = [\boldsymbol{\beta}^{*}, \sigma_{Y}, \boldsymbol{\sigma}_{\mathbf{X}}, \mathit{vecp}(\boldsymbol{P}_{\mathbf{X}})],$$

where $\mathit{vecp}(\bullet)$ is the matrix operator extracting lower-diagonal elements from a symmetric matrix into a vector.

SEM for Semipartial Correlation Coefficients

- Using the equivalent constraint function, semipartial correlation coefficients can be obtained as:

$$\boldsymbol{\Sigma}_{Y\mathbf{X}} = \mathbf{D}_{\sigma} \left[\begin{array}{c|c} \boldsymbol{\beta}^{*\prime} \boldsymbol{P}_{\mathbf{X}} \boldsymbol{\beta}^{*} + \sigma_{\varepsilon}^{2} & \boldsymbol{\beta}^{*\prime} \boldsymbol{P}_{\mathbf{X}} \\ \hline \boldsymbol{P}_{\mathbf{X}} \boldsymbol{\beta}^{*} & \boldsymbol{P}_{\mathbf{X}} \end{array} \right] \mathbf{D}_{\sigma}$$

  where standardized regression coefficients are computed by the user-defined parameter function:

$$\boldsymbol{\beta}^{*} = \boldsymbol{\rho}_{sr} \times (\mathbf{1} - \mathit{Diag}[\boldsymbol{P}_{\mathbf{X}}^{-1}])^{-0.5}$$

- The set of freely estimated model parameters equals:

$$\boldsymbol{\theta}_{sr} = [\boldsymbol{\rho}_{sr}, \sigma_{Y}, \boldsymbol{\sigma}_{\mathbf{X}}, \mathit{vecp}(\boldsymbol{P}_{\mathbf{X}})],$$

  with the same constraint function on σ_{ε}^{2} being imposed.

SEM for Improvement in R-squared

- Using the equivalent constraint function, improvement in R-squared values can be obtained as:

$$\boldsymbol{\Sigma}_{Y\mathbf{X}} = \mathbf{D}_{\sigma} \left[\begin{array}{c|c} \boldsymbol{\beta}^{*\prime} \boldsymbol{P}_{\mathbf{X}} \, \boldsymbol{\beta}^{*} + \sigma_{\varepsilon}^{2} & \boldsymbol{\beta}^{*\prime} \boldsymbol{P}_{\mathbf{X}} \\ \hline \boldsymbol{P}_{\mathbf{X}} \, \boldsymbol{\beta}^{*} & \boldsymbol{P}_{\mathbf{X}} \end{array} \right] \mathbf{D}_{\sigma}$$

where standardized regression coefficients are instead computed by the user-defined parameter function:

$$\boldsymbol{\beta}^{*} = \boldsymbol{\kappa} [\boldsymbol{\Delta}_{R^2} \times (\mathbf{1} - \mathit{Diag}[\boldsymbol{P}_{\mathbf{X}}^{-1}])^{-0.5}]$$

and where $\boldsymbol{\kappa}$ is a sign vector with values equal to $+1$ or -1.

- The set of freely estimated model parameters equals:

$$\boldsymbol{\theta}_{\Delta_{R^2}} = [\boldsymbol{\Delta}_{R^2}, \sigma_{Y}, \boldsymbol{\sigma}_{\mathbf{X}}, \mathit{vecp}(\boldsymbol{P}_{\mathbf{X}})],$$

with the same constraint function on σ_{ε}^{2} once again being imposed.

Standard Errors for Regression Coefficients

- Covariance matrix of regression coefficient parameters under OLS equals (where $\tilde{\mathbf{X}} = [\mathbf{1}|\mathbf{X}]$):

$$Cov(\hat{\boldsymbol{\beta}})_{OLS} = (N-J)^{-1} \sum_{n=1}^{N} e_n^2 \left(\tilde{\mathbf{X}}\tilde{\mathbf{X}}'\right)^{-1}$$

- White (1980) proposed a heteroscedastic-consistent estimator when the regression model is misspecified:

$$Cov(\hat{\boldsymbol{\beta}})_{HC0} = \left(\tilde{\mathbf{X}}\tilde{\mathbf{X}}'\right)^{-1} \left(\tilde{\mathbf{X}}' \, Diag\left[e_n^2\right] \, \tilde{\mathbf{X}}\right) \left(\tilde{\mathbf{X}}\tilde{\mathbf{X}}'\right)^{-1}$$

- Hinkley (1977) had earlier and independently proposed a finite sample-adjusted version of HC0 that is less biased:

$$Cov(\hat{\boldsymbol{\beta}})_{HC1} = \frac{N}{N-J} \left(\tilde{\mathbf{X}}\tilde{\mathbf{X}}'\right)^{-1} \left(\tilde{\mathbf{X}}' \, Diag\left[e_n^2\right] \, \tilde{\mathbf{X}}\right) \left(\tilde{\mathbf{X}}\tilde{\mathbf{X}}'\right)^{-1}$$

Two Further Developments

- MacKinnon & White (1985) subsequently proposed two other HC-based estimators to reduce bias further.
- Let $h_n = \mathbf{x}_n' \left(\mathbf{X}\mathbf{X}'\right)^{-1} \mathbf{x}_n$ signify a person's leverage statistic.

$$Cov(\hat{\boldsymbol{\beta}})_{HC2} = \left(\tilde{\mathbf{X}}\tilde{\mathbf{X}}'\right)^{-1} \left(\tilde{\mathbf{X}}' \, Diag\left[\frac{e_n^2}{(1-h_n)}\right] \tilde{\mathbf{X}}\right) \left(\tilde{\mathbf{X}}\tilde{\mathbf{X}}'\right)^{-1}$$

$$Cov(\hat{\boldsymbol{\beta}})_{HC3} = \left(\tilde{\mathbf{X}}\tilde{\mathbf{X}}'\right)^{-1} \left(\tilde{\mathbf{X}}' \, Diag\left[\frac{e_n^2}{(1-h_n)^2}\right] \tilde{\mathbf{X}}\right) \left(\tilde{\mathbf{X}}\tilde{\mathbf{X}}'\right)^{-1}$$

- Long & Ervin (2000) recommended HC3 as the best general estimator among heteroscedastic-consistent ones.
- All these estimators use raw sample data.

Normal Standard Errors in SEM

- Under multivariate normality, the covariance matrix of model parameters using ML estimation equals:

$$Cov(\hat{\boldsymbol{\theta}})_{NT} = N^{-1} \left(\boldsymbol{\Delta}_{\theta}' \boldsymbol{\Omega} \boldsymbol{\Delta}_{\theta} \right)^{-1}$$

where $\boldsymbol{\Omega}$ is the asymptotic covariance matrix of population covariances for both the dependent and J independent variables.

$$\boldsymbol{\Omega} = \left[2 \times \mathbf{K}_p' \left(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} \right) \mathbf{K}_p \right]^{-1}$$

and $\mathbf{K}_p$ is a transition matrix.

- For linear regression:

$$Cov(\hat{\boldsymbol{\theta}})_{NT} \equiv Cov(\hat{\boldsymbol{\beta}})_{OLS}.$$

Robust Standard Errors in SEM

- SEs based on $Cov(\hat{\boldsymbol{\theta}})_{NT}$ are too small if data are non-normal. A more robust set can be derived using the sandwich estimator:

$$Cov(\hat{\boldsymbol{\theta}})_{R0} = N^{-1} \left(\boldsymbol{\Delta}_{\theta}' \boldsymbol{\Omega} \boldsymbol{\Delta}_{\theta}\right)^{-1} \left(\boldsymbol{\Delta}_{\theta}' \, \boldsymbol{\Omega} \, \tilde{\boldsymbol{\Omega}} \, \boldsymbol{\Omega} \, \boldsymbol{\Delta}_{\theta}\right) \left(\boldsymbol{\Delta}_{\theta}' \boldsymbol{\Omega} \boldsymbol{\Delta}_{\theta}\right)^{-1}$$

- Yuan & Hayashi (2006) show the correct specification of $\tilde{\boldsymbol{\Omega}}$ is in general given by:

$$\tilde{\boldsymbol{\Omega}} = N^{-1} \sum_{n=1}^{N} vech\left[\tilde{\boldsymbol{\Sigma}}_n\right] vech\left[\tilde{\boldsymbol{\Sigma}}_n\right]'$$

where

$$\tilde{\boldsymbol{\Sigma}}_n = \left(\mathbf{X}_n - \bar{\mathbf{X}}\right)\left(\mathbf{X}_n - \bar{\mathbf{X}}\right)' - \boldsymbol{\Sigma}$$

- For linear regression, it can be shown that:

$$Cov(\hat{\boldsymbol{\theta}})_{R0} \equiv Cov(\hat{\boldsymbol{\beta}})_{HC0}.$$

Extending Robust SEs to the HC3 Estimator

- Robust SEs in SEM do not currently cover MacKinnon & White's (1985) improved heteroscedastic-consistent estimators.

- Performing linear regression within a SEM framework enables this to be accomplished as follows:

$$e_n^2 \equiv vech\left[\tilde{\Sigma}_n\right] vech\left[\tilde{\Sigma}_n\right]'$$

Therefore, we can define a more robust estimator of the asymptotic covariance matrix of population covariance as:

$$\tilde{\Omega}^* = (1/N) \sum_{n=1}^{N} vech\left[\tilde{\Sigma}_n\right] (1 - h_n)^{-2} vech\left[\tilde{\Sigma}_n\right]'$$

which results in

$$Cov(\hat{\theta})_{R3} = N^{-1} \left(\Delta_\theta' \Omega \Delta_\theta\right)^{-1} \left(\Delta_\theta' \Omega \tilde{\Omega}^* \Omega \Delta_\theta\right) \left(\Delta_\theta' \Omega \Delta_\theta\right)^{-1}$$

Extending Robust SEs to the HC3 Estimator (cont.)

- This result means that:

$$Cov(\hat{\boldsymbol{\theta}})_{R3} \equiv Cov(\hat{\boldsymbol{\beta}})_{HC3}.$$

- The SEM-based framework, however, is far more flexible and straightforward.
- Two steps are required:
  1. Select one of $\boldsymbol{\theta}_{\beta}$ or $\boldsymbol{\theta}^{*}_{\beta}$ or $\boldsymbol{\theta}_{sr}$ or $\boldsymbol{\theta}_{\Delta_{R^2}}$ for $\hat{\boldsymbol{\theta}}$ in $Cov(\hat{\boldsymbol{\theta}})_{R3}$.
  2. Calculate the appropriate Jacobian matrix $\boldsymbol{\Delta}_{\theta}$ for the chosen parameter vector.
- The second step can be easily implemented using automatic differentiation algorithms.

Investigating Confidence Interval Coverage

1. Used a $3 \times 5 \times 4 \times 4 \times 4$ experimental design for simulation.

1. Number of IVs	$\{2, 5, 8\}$
2. Semipartial Correlation	$\{0, .05, .15, .35, .55\}$
3. Sample size	$\{50, 100, 300, 1000\}$
4. Interim R-squared Value	$\{0, .10, .20, .50\}$
5. Tolerance of IV	$\{0.20, 0.50, 0.80, 1.00\}$

2. 10,000 replications in each of the 960 cells of the design.
3. Data for all variables had excess kurtosis equal to 7.0.
4. Generated using methods proposed by Headrick & Kowalchuck (2008).
5. Calculated proportion of replications in which 95% confidence intervals captured the designated population effect size value.
6. Two outcome measures: *accuracy* and *robustness*.

Investigating Confidence Interval Coverage (Cont.)

7. **Accuracy** was measured by calculating mean square error (MSE), which can itself be further decomposed into (squared) *bias* and *imprecision*:

$$\mathcal{E}\left[\left(\hat{\theta}-\theta\right)^2\right] = \left(\mathcal{E}\left[\hat{\theta}-\theta\right]\right)^2 + \mathcal{E}\left[\left(\hat{\theta}-\mathcal{E}[\hat{\theta}]\right)^2\right]$$

8. **Robustness** was assessed by Serlin's (2000) range null hypothesis test, using Bradley's liberal criterion (i.e., $\pm$ 2.5%).

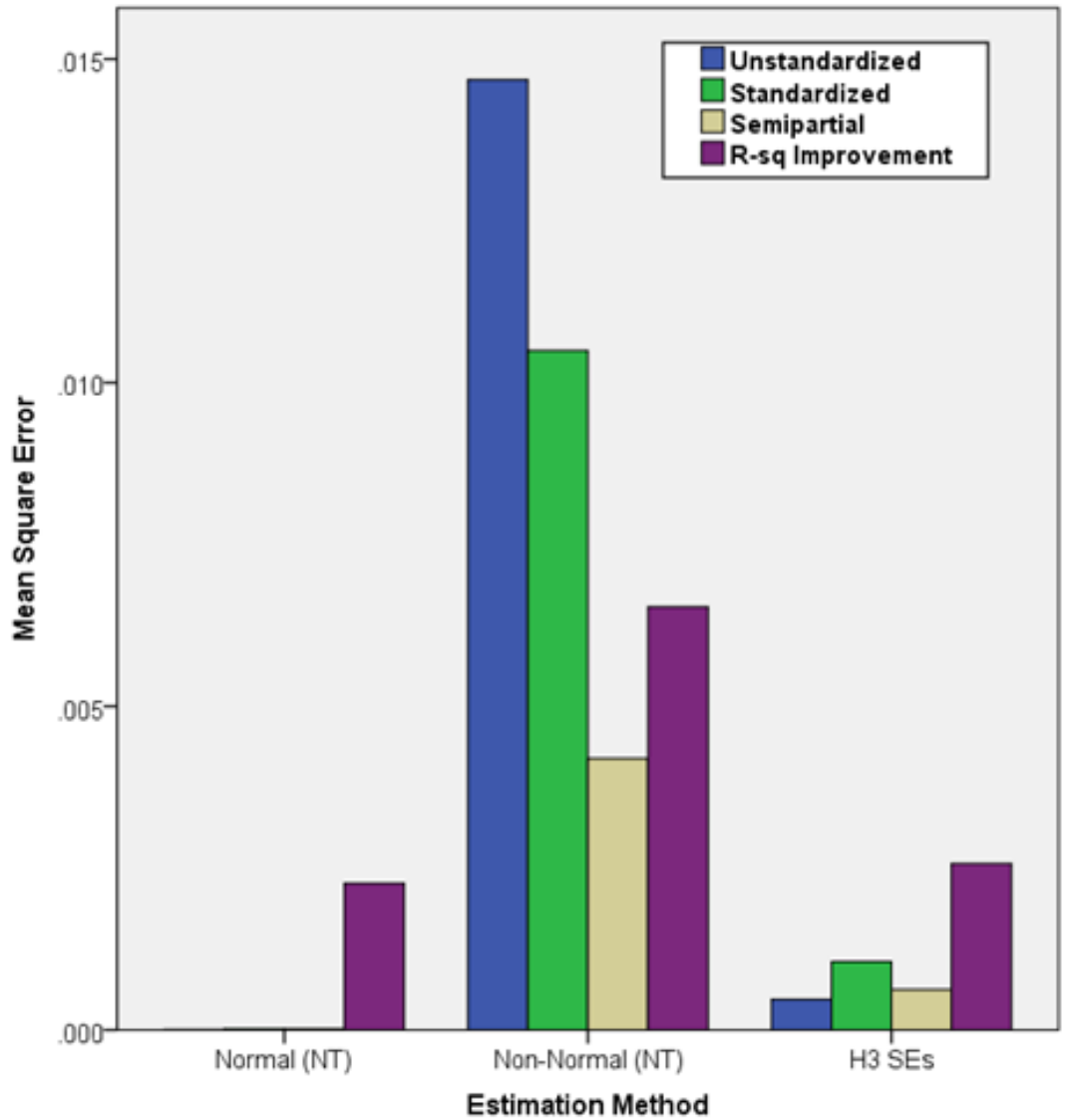
9. The resultant measure was the proportion of replications for which the Serlin's null hypothesis of non-robustness was rejected.

10. Measures of *accuracy* and *robustness* are complimentary, in that the most accurate interval may not necessarily be robust whereas a robust interval may not be the most accurate.

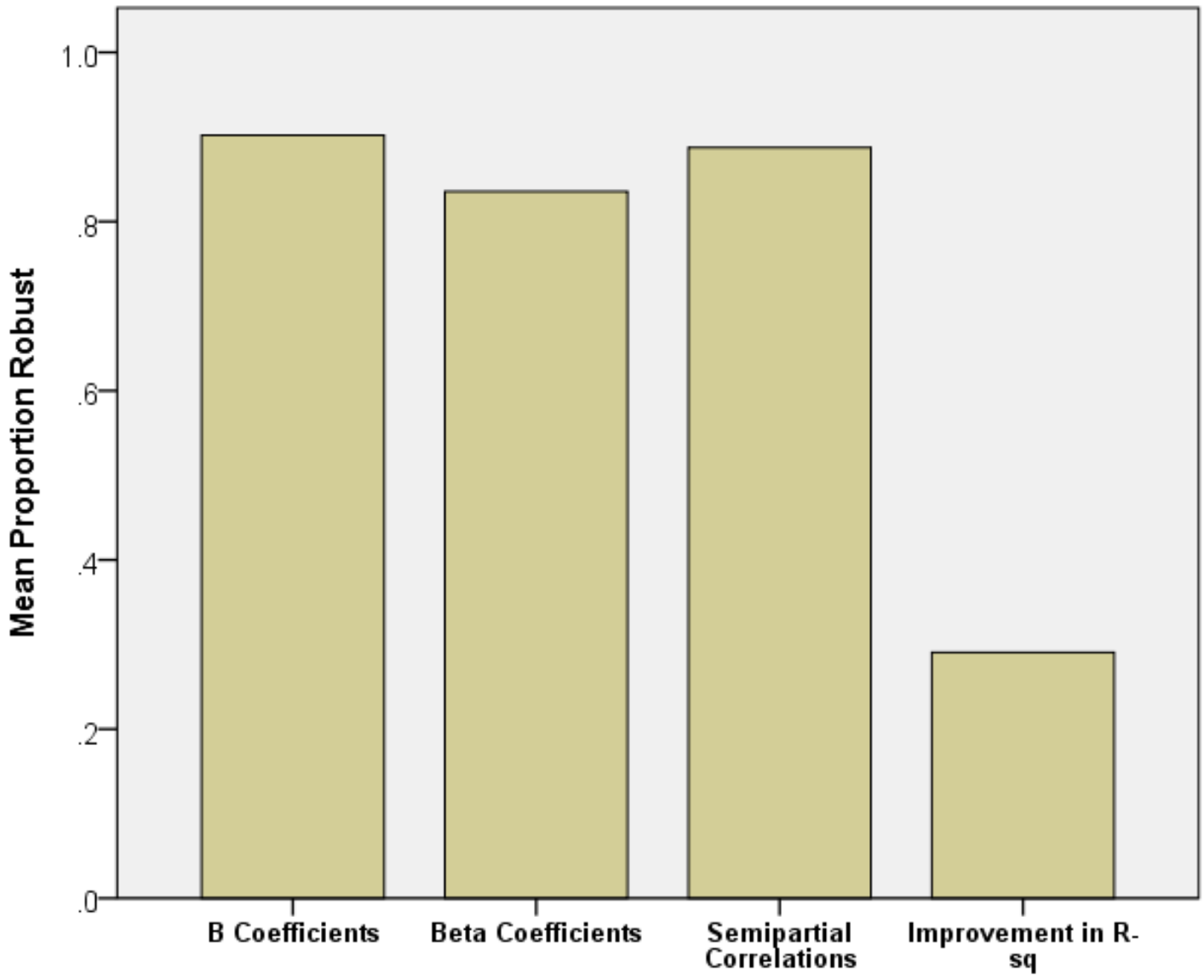
Brief Explanation about Confidence intervals

- All 95% confidence intervals were calculated using the relevant *t* distribution for better small sample performance.
- Confidence intervals for *semipartial correlation* and *improvement in R-squared* were calculated using a method proposed in Browne (1982).
- General approach is to:
  (a) apply an appropriate transformation on the bounded parameter to make it unbounded,
  (b) construct the symmetric confidence interval in usual way, and then
  (c) invert the transformation on both limits to obtain a bounded, asymmetric interval
- All calculations and simulation was undertaken in MATLAB.

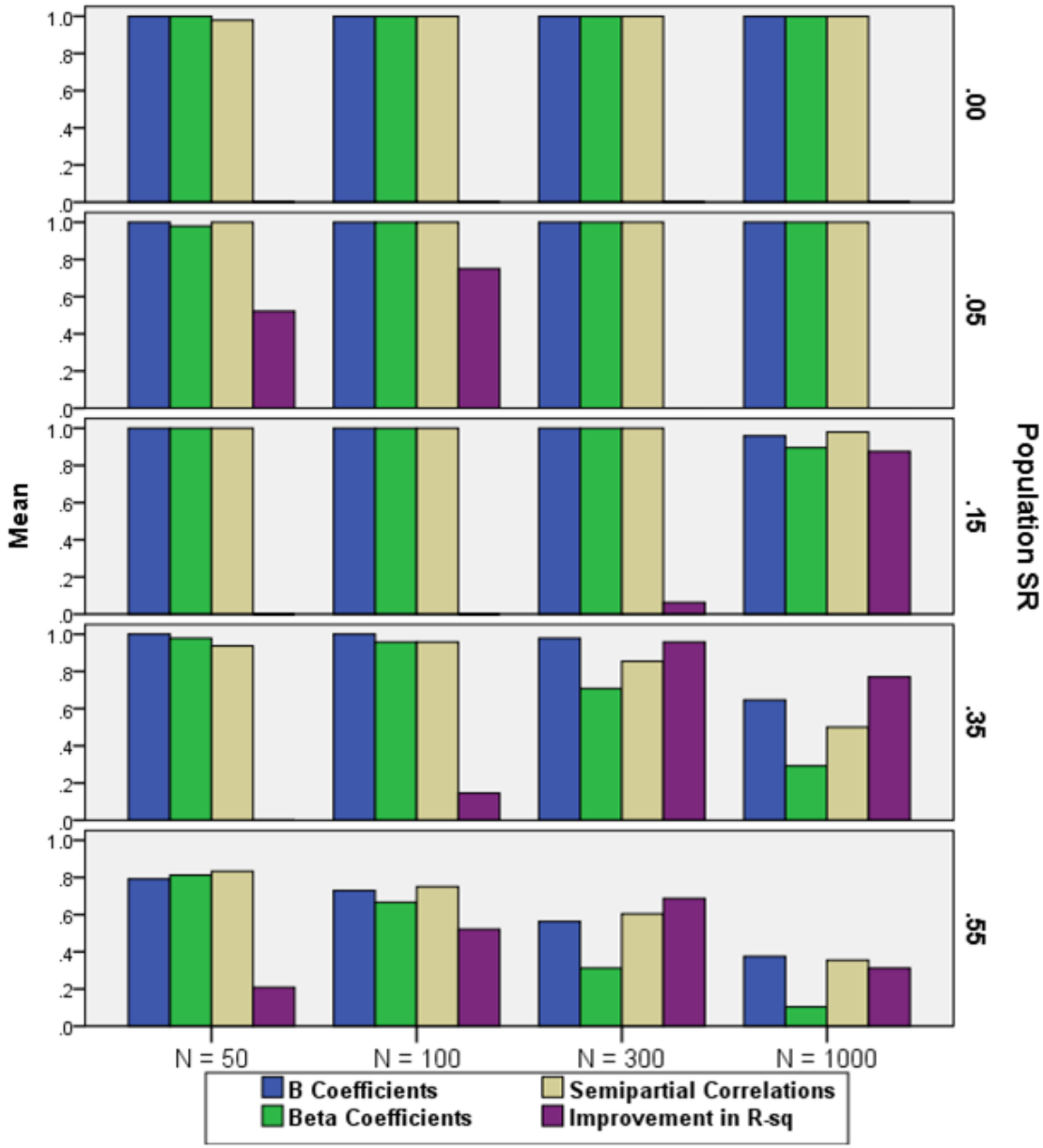
Accuracy of Confidence Intervals



Robustness of Confidence Intervals



Robustness of Confidence Intervals (cont.)



Summary and Further Potentially Developments

- The robust SEs work far better than normal ones under non-normality for 2 of 3 effect sizes under the majority of conditions.
- Preliminary evidence that the same applies under heteroscedasticity.
- General SEM framework is quite flexible to include other effect sizes such as model R-square and (potentially) relative importance indices.
- Other estimators, such as Yuan & Benter's (1999) robust covariance methods using Huber weights, could also be investigated.

And Finally...

Thank You for Coming and Listening.

Any Questions?