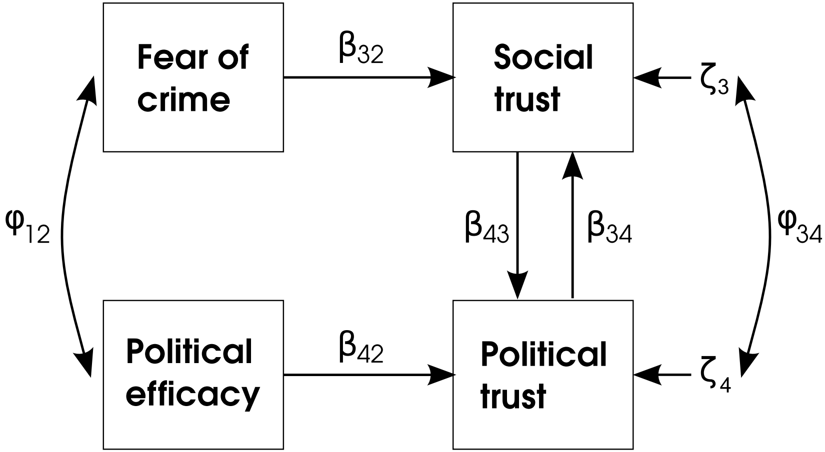
Evaluating the conditional relative effects of TSE components on the variance of parameter estimates in structural equation models (SEM)

Daniel Oberski
Willem Saris

TILBURG UNIVERSITY

UNIVERSITAT POMPEU FABRA

Example structural equation model (Saris & Gallhofer 2007)



Parameters to estimate: $\beta_{32}, \beta_{42}, \beta_{43}, \beta_{34}, \phi_{12}, \phi_{34}, \phi_{11}, \phi_{22}, \phi_{33}, \phi_{44}$.

TSE components' effect on structural equation models

TSE components' effect on structural equation models

Survey errors may affect:

- Bias in SEM parameter estimates;
- Variance of SEM parameter estimates.

We focus on **variance**:

- Few empirical studies finding bias in regression-type coefficients due to survey errors (though can't rule it out);
- Relatively little known about TSE effect on variance of regression-type coefficients and SEM (also remarked by Groves in his 2011 Morris Hansen lecture).

Three TSE components' effect on SEM models

- **Measurement error** :
  - causes bias when not corrected for (e.g. Fuller 1987);
  - increases variance when corrected for;
  - increases variance even more when error is estimated rather than known (Oberski & Satorra 2012)
- **Unequal sampling probabilities and clustering**:
  - Different estimator needed for covariance matrix, need to take into account.
  - Complex sample covariance matrix of observed variables will itself have different asymptotic variance matrix, $\mathbf{\Gamma}$.
- **Non-normality of observed variables**:
  - Does not affect consistency, but affects asymptotic variance matrix $\mathbf{\Gamma}$ of the covariances between the observed variables.
  - Especially multivariate **kurtosis** influences the variance.

Evaluating the effect of TSE components through model-based variance estimation in structural equation models

SEM: model-based variance estimation

Under the model, asymptotic variance of the estimates $\hat{\theta}$ is

$$avar(\hat{\boldsymbol{\theta}}) = n^{-1}(\boldsymbol{\Delta}'\boldsymbol{V}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\boldsymbol{V}\boldsymbol{\Gamma}\boldsymbol{V}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\boldsymbol{V}\boldsymbol{\Delta})^{-1}, \qquad (1)$$

where (Satorra 1989):

- $\boldsymbol{\Delta}$ is first derivative of implied covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ w.r.t. the parameters $\boldsymbol{\theta}$ (given by Neudecker & Satorra, 1991);
- $\boldsymbol{\Gamma}$ is fourth-order moments of the observed variables;
- $\boldsymbol{V}$ is second derivative w.r.t. the parameters of the fitting function (usually Maximum Likelihood).

Consistent estimate of $avar(\hat{\boldsymbol{\theta}})$ by replacing $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}$ in expressions for $\boldsymbol{\Delta}, \boldsymbol{\Gamma}$, and $\boldsymbol{V}$.

$$avar(\hat{\boldsymbol{\theta}}) = n^{-1}(\boldsymbol{\Delta}'\boldsymbol{V}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\boldsymbol{V}\boldsymbol{\Gamma}\boldsymbol{V}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\boldsymbol{V}\boldsymbol{\Delta})^{-1}.$$

Depending on whether error components are taken into account or not,

- Estimator for θ may be different, leading to changes in $\mathbf{V}$
- Asymptotic variance matrix of the observed (co)variances may be different, leading to changes in $\boldsymbol{\Gamma}$

Conditional effects

Conditional effect of survey error components on variance of $\hat{\boldsymbol{\theta}}$ can be evaluated by evaluating $avar(\hat{\boldsymbol{\theta}})$ under different conditions.

Can show:

- Determinant of $\mathbf{V}$ a decreasing function of the reliability;
- Determinant of $\mathbf{\Gamma}$ increasing function of the kurtosis;
- Determinant of $\mathbf{\Gamma}$ increasing function of the icc (Scott & Holt 1982).

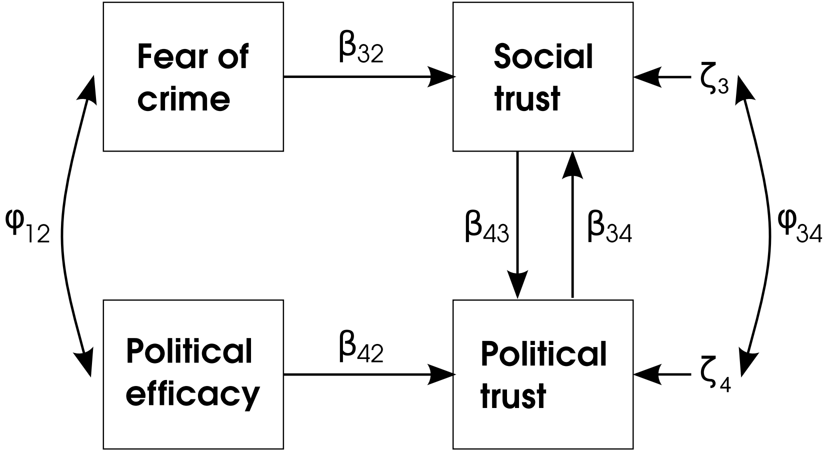
This motivates an indicator of the conditional multiplicative effect of measurement error, nonnormality, and clustering on *avar*$(\boldsymbol{\theta})$.

Conditional relative efficiency

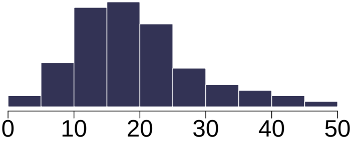
Conditional relative efficiency due to survey error source defined as the variance of $\hat{\boldsymbol{\theta}}$ taking into account survey error source divided by normal-theory variance under simple random sample and fixed measurement error.

Example: evaluation of survey errors' conditional effects on variance of SEM parameter estimates

Example structural equation model (Saris & Gallhofer 2007)



Example SEM - data (http://www.europeansocialsurvey.org/)

Survey:	European Social Survey, Round 4
Fieldwork period:	01 Sept 2008 -- 11 Jan 2009
Country:	Denmark (Danish) 
Conducted by:	SFI-SURVEY
Survey mode:	CAPI
Sampling design:	SRS of individuals age 15+ from Danish Central Person Register
Achieved sample size:	1610
Response rate:	54% (only 0.7% ineligibles)
Number of interviewers:	80
Number of interviews/intvr:	

Example SEM - descriptives

Variable	socialTrust	systemTrust	fearCrime	efficacy
Mean	20	21	5	7
Stddev	4.7	4.8	1.7	1.7
Skewness	-0.7	-0.79	0.77	-0.25
Kurtosis	-0.8	-0.54	-0.31	0.05
icc (intvr)	0.25	0.11	0.17	0.11
$\hat{\rho}_X$(s.e.)	0.73(0.01)	0.77(0.01)	0.57(0.02)	0.64(0.03)
$\hat{\psi}$(s.e.)	6.0(0.22)	6.3(0.24)	1.3(0.04)	1.2(0.07)

Conditional relative efficiency due to measurement error, non-normality, and interviewer clustering

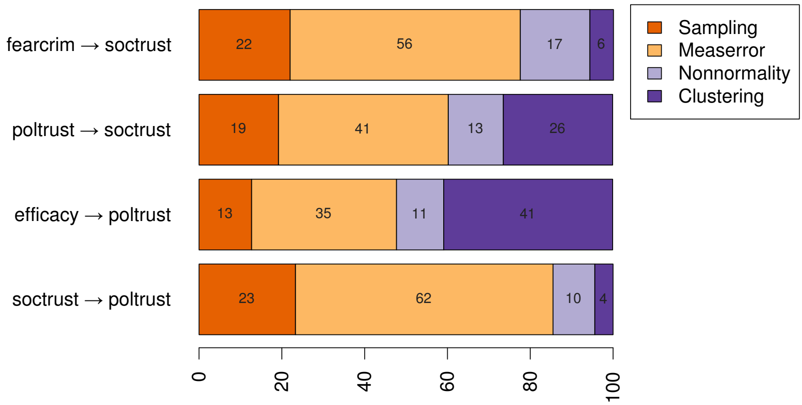
Parameter estimates & square root conditional relative efficiency (creff)

Parameter	$\hat{\theta}$	$\text{creff}_{\text{measerr}}$	$\text{creff}_{\text{non-normality}}$	$\text{creff}_{\text{clustering}}$
soctrust $\rightarrow$ poltrust	0.77	1.92	1.06	1.02
efficacy $\rightarrow$ poltrust	0.51	1.94	1.11	1.30
poltrust $\rightarrow$ soctrust	0.30	1.77	1.10	1.17
fearcrim $\rightarrow$ soctrust	-0.68	1.88	1.10	1.03
ϕ(poltrust, soctrust)	6.54	1.98	1.11	1.19
ϕ(efficacy)	1.64	1.71	0.99	1.11
ϕ(fearcrime)	1.71	1.78	1.07	1.16
cov(efficacy, fearcrime)	-0.60	1.73	0.77	1.10

- **Random measurement error** :
- Causes between 71% and 98% increase in s.e.'s here
- Effect of **interviewer clustering** is less than the effect of measurement error but also considerable:
- creff ranging between 1.03 and 1.30 ;
- **Non-normality** mostly increases the variance of the estimates -- except cov(efficacy, fearcrime)
- % increase in s.e. due to non-normality is modest compared with effects of interviewer clustering and measurement error.

Conditional decomposition of variance for four regression parameters

Percentage of variance of the regression coefficients of the model contributed by each error component.



Conclusions example

- In the example, measurement error, when corrected for, provided a large contribution to the MSE;
- But when not corrected for the MSE was even higher;
- This suggests that in this particular study, it has been cost-effective to include multiple indicators of the different concepts, allowing for correction for measurement error;
- There is also a limit, however, to the number of indicators that it is cost-effective to include as other survey error components gain in relative importance;
- It is not clear whether this occurs in other studies as well -- could be evaluated.

General conclusions

- SEM parameters generalize second-order statistics such as subclass mean differences, regression coefficients, factor analyses, etc. -- often of interest to survey users;
- Developed a conditional measure of the relative contribution of different survey error components to variance of SEM parameters;
- May be used for evaluating different aspects of survey design simultaneously in general context;
- Precludes the need for extensive simulations.

Disadvantage:

- Measures are necessarily dependent on parameter values and other design features.

Thank you for your attention!

Daniel Oberski
doberski@uvt.nl

Appendix

Different survey error components' effect on estimator

	Distribution	Clustering/ weighting	Measurement error	Γ	$\hat{\Sigma}_\eta$	Σ_Ψ
1	Normal	-	-	$\hat{\Gamma}^*$	S	0
2	Normal	-	Fixed	$\hat{\Gamma}^*$	$B^{-1}\Phi B^{-1}$	0
	Normal	-	Estimated	$\hat{\Gamma}^*$	$B^{-1}\Phi B^{-1}$	$\hat{\Sigma}_\Psi$
	Normal	Yes	-	$\hat{\Gamma}^{(c)*}$	$S^{(c)}$	0
	Normal	Yes	Fixed	$\hat{\Gamma}^{(c)*}$	$B^{-1}\Phi B^{-1}$	0
	Normal	Yes	Estimated	$\hat{\Gamma}^{(c)*}$	$B^{-1}\Phi B^{-1}$	$\hat{\Sigma}_\Psi$
	Non-normal	-	-	$\hat{\Gamma}$	S	0
3	Non-normal	-	Fixed	$\hat{\Gamma}$	$B^{-1}\Phi B^{-1}$	0
	Non-normal	-	Estimated	$\hat{\Gamma}$	$B^{-1}\Phi B^{-1}$	$\hat{\Sigma}_\Psi$
	Non-normal	Yes	-	$\hat{\Gamma}^{(c)}$	$S^{(c)}$	0
4	Non-normal	Yes	Fixed	$\hat{\Gamma}^{(c)}$	$B^{-1}\Phi B^{-1}$	0
	Non-normal	Yes	Estimated	$\hat{\Gamma}^{(c)}$	$B^{-1}\Phi B^{-1}$	$\hat{\Sigma}_\Psi$

Parameter estimates and standard errors under different conditions

Parameter	$\hat{\theta}$	$\sigma_1(\hat{\theta})$	$\sigma_2(\hat{\theta})$	$\sigma_3(\hat{\theta})$	$\sigma_4(\hat{\theta})$
soctrust $\rightarrow$ poltrust	0.77	0.08	0.16	0.17	0.17
efficacy $\rightarrow$ poltrust	0.51	0.08	0.16	0.18	0.23
poltrust $\rightarrow$ soctrust	0.30	0.11	0.19	0.21	0.25
fearcrim $\rightarrow$ soctrust	-0.68	0.12	0.22	0.25	0.25
ϕ(poltrust, soctrust)	6.54	1.60	3.17	3.52	4.20
ϕ(efficacy)	1.64	0.06	0.10	0.10	0.11
ϕ(fearcrime)	1.71	0.06	0.11	0.12	0.14
cov(efficacy, fearcrime)	-0.60	0.06	0.10	0.08	0.09

Correct for random measurement error or not?

- Correcting for measurement error will increase the estimated variance of the parameters;
- But if the model is not corrected for measurement error the parameter estimates of interest will be biased.
- Which is worse?
- The Mean Square Error (MSE) of the naive (not corrected for measurement error) estimate then equals bias$^2 + \sigma^2(\theta_{\text{naive}})$.
- Calculate approximate percentage of mean square error in the naive estimates due to the bias and the variance, respectively.

% of MSE in naive estimates due to bias and variance

	soctrust → poltrust	poltrust → soctrust	efficacy → poltrust	fearcrim → soctrust
$\hat{\theta}_{\text{correct}}$	0.77	0.3	0.51	-0.68
$\hat{\theta}_{\text{naive}}$	0.83	0.44	0.28	-0.31
$\widehat{\text{bias}}^2$	0.0036	0.0196	0.0529	0.1369
$\sigma^2(\hat{\theta}_{\text{naive}})$	0.02	0.04	0.02	0.01
$\sqrt{\text{MSE}}$	0.16	0.25	0.26	0.38
σ_m	0.15	0.16	0.15	0.23
%MSE,bias	15%	32%	76%	89%
%MSE,var.	85%	68%	23%	10%