SIMPLE CORRECTION FOR MEASUREMENT ERRORS WITH STATA

8ª Reunión Usuarios Stata, Madrid
22th October 2015
Anna DeCastellarnau
ESS-CST, Universitat Pompeu Fabra
anna.decastellarnau@upf.edu

"A simple procedure to correct for measurement errors in survey research"

Written by: Anna DeCastellarnau and Willem Saris

http://essedunet.nsd.uib.no/cms/topics/measurement/

	Results without corrections	Results with corrections	
Effects	**Regression coefficients**	**Regression coefficients**	
Dependent V1<-			
V2	0.248**	0.406**	+0.158
V3	-0.022	0.039	+0.061
V4	0.246**	0.415**	+0.169
V5	0.215**	0.103**	-0.112
V6	-0.066**	-0.150**	+0.084
R^2	**0.226 (22.6%)**	**0.456 (45.6%)**	

** if α<1% and * if 1%<α<5%

+0.23

- Increase in effects of more than 1 point on average
- Even changes in the sign of the effect happen
- Increase in more than factor 2 in the explained variance

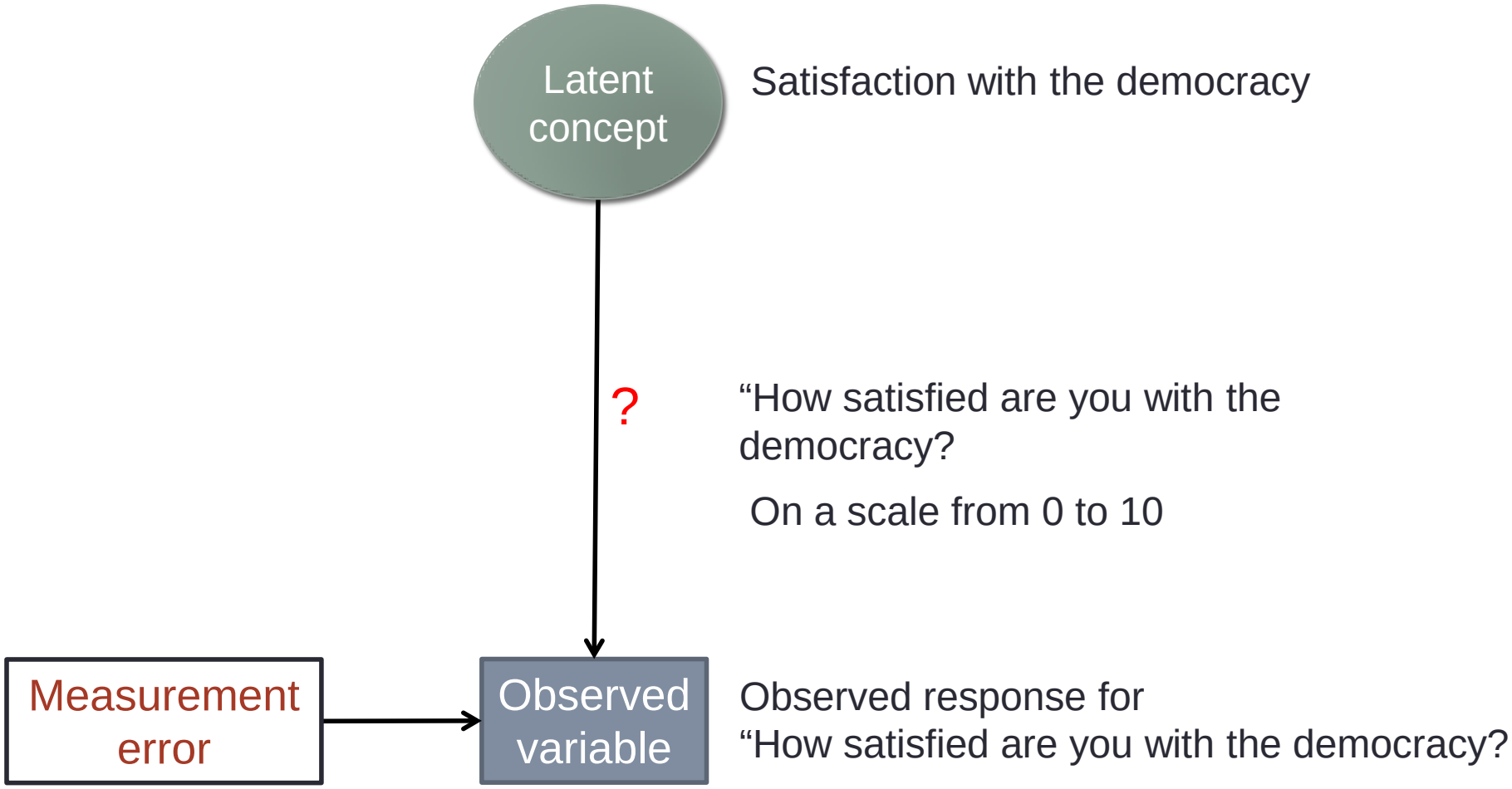
OUTLINE

Theory

Applicability using Stata

Benefits and possibilities

WHAT DO WE MEASURE?



WHAT IS MEASUREMENT ERROR?

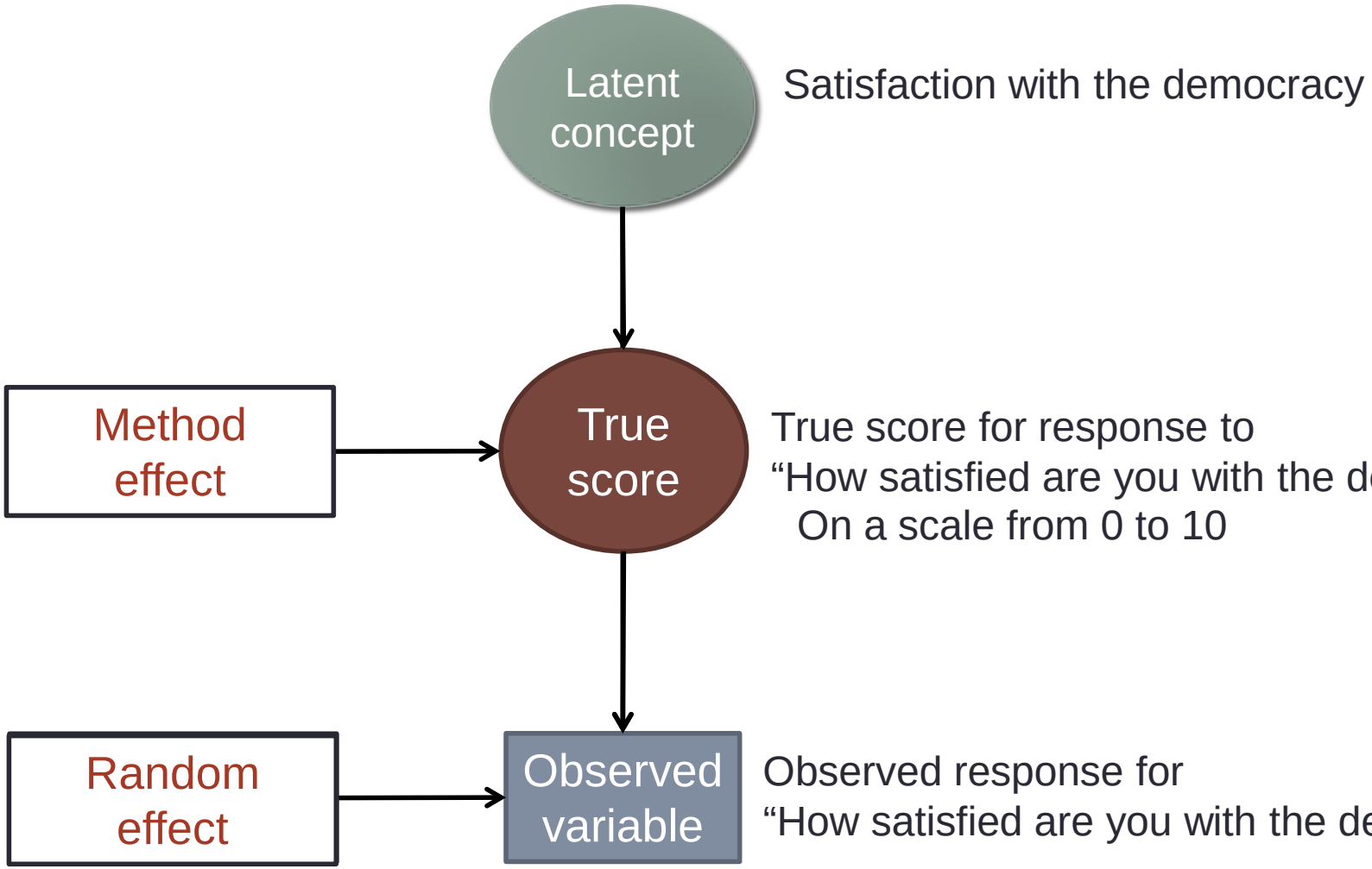
- There are two components of M.E.:

- **Random error**
  - Captures the effect of unintended and unpredictable fluctuations of the respondents, interviewers, coders, etc…

- **Systematic error or method effect**
  - Captures the effect of the reaction of the respondents to a particular formulation of a question.
  - Respondents can react differently to different formulations of questions even if the concept asked is not changed.

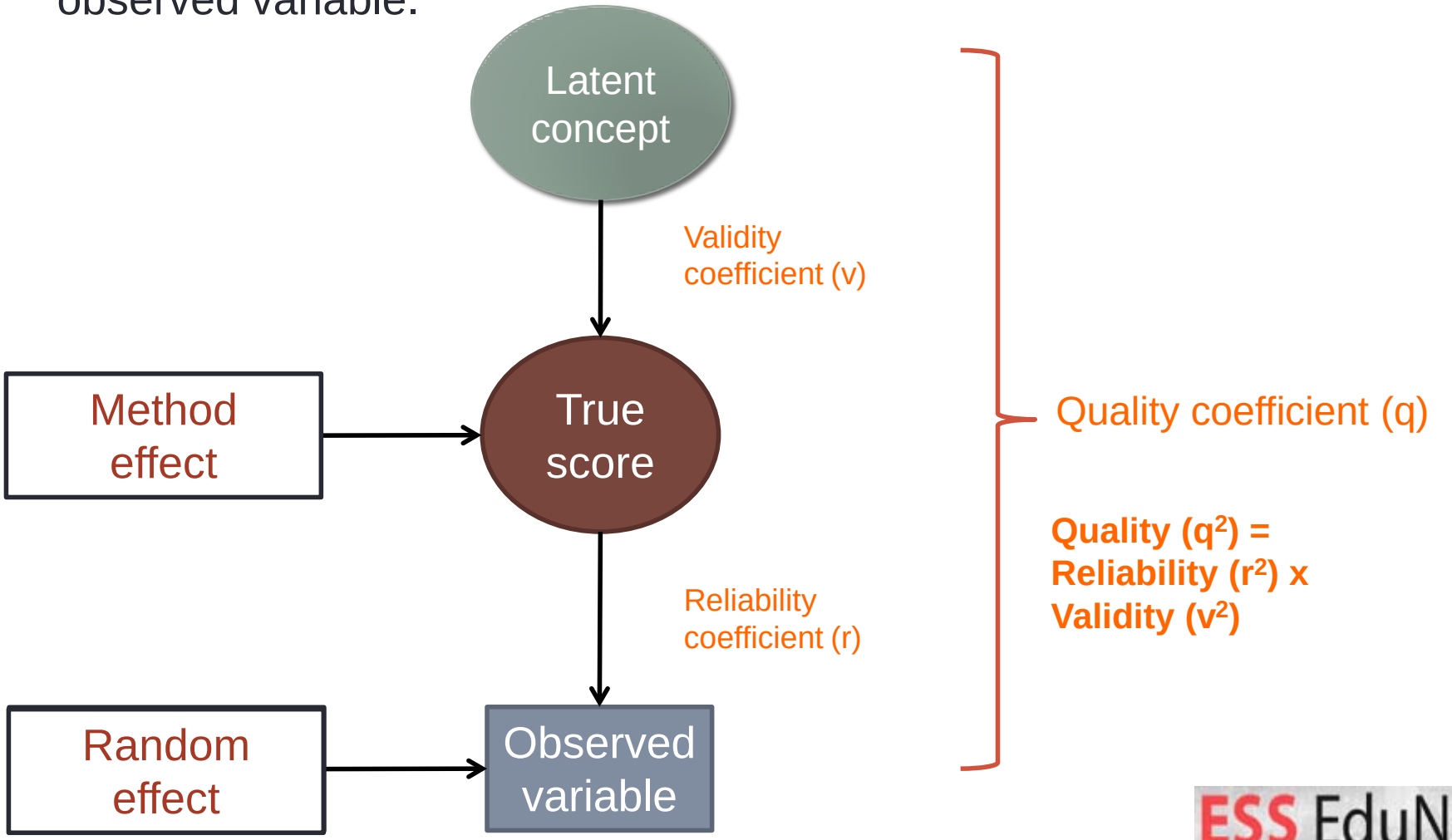
Already discussed in: Campbell and Fiske (1959), Schuman and Presser (1981) and Sudman and Bradburn (1982) and many others.

WHAT DO WE MEASURE? (II)



HOW IS THE QUALITY DEFINED?

- **Quality (q^2)** is the strength between the latent concept and the observed variable.



HOW DO WE OBTAIN QUALITY?

- **Option 1: Conduct a Multitrait-Multimethod (MTMM) experiment.**

- **Option 2: Use alternative approach…**
  - **Over the last decades many MTMM data have been collected**
    - Database of:
      - 3,726 questions with quality information
      - In more than 20 countries and languages
      - From multiple surveys

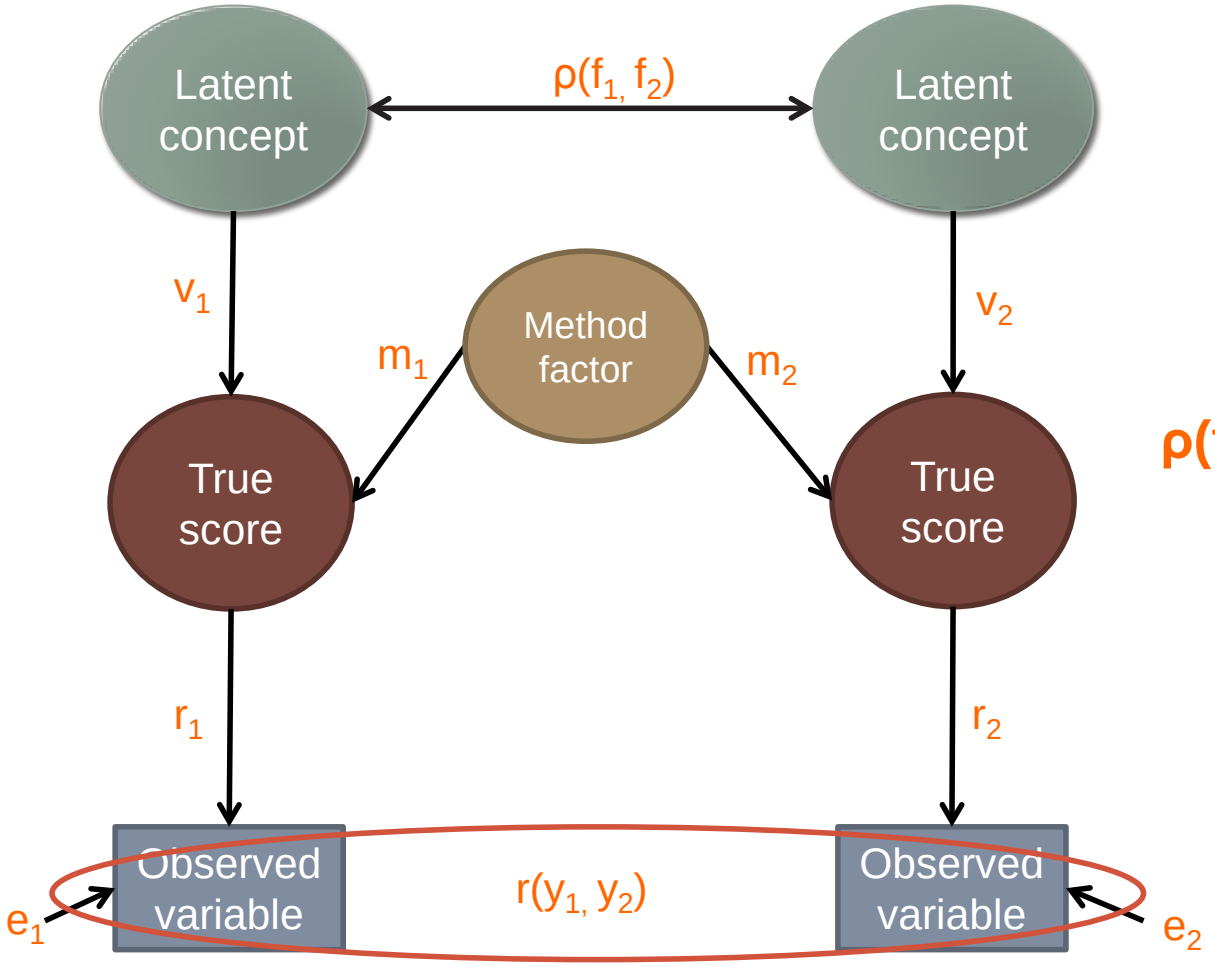
  - **The formal and linguistic characteristics of these questions were carefully coded**
    - The quality obtained from the MTMM experiments could be related to the characteristics of the survey questions.

  - **A new tool was developed:**
    - Allows to predict the quality of survey questions
    - Requires only the coding of the characteristics of the survey question
    - Provides the information about the reliability and validity
    - It is available online for free: [sqp.upf.edu](sqp.upf.edu)

Already discussed in: Saris and Gallhofer (2014), Oberski et al (2011).

HOW CAN WE SIMPLY CORRECT FOR M.E.?

- **Correction of the observed correlation matrix**



- **Formula:**

$$r(y_1, y_2) = r_1 v_1 \rho(f_1, f_2) v_2 r_2 + r_1 m_1 m_2 r_2$$

$$\rho(f_1, f_2) = \frac{[r(y_1, y_2) - CMV_{12}]}{q_1 q_2}$$

EXAMINING THE FORMULA

$$\rho(f_1, f_2) = \frac{[r(y_1, y_2) - CMV_{12}]}{q_1 q_2}$$

- The correlation between two observed variables **$r(y_1, y_2)$** is known.
- The common method variance (CMV) is the factor that decreases the over estimation of the observed correlation of those variables that share the same method.
- The CMV between two variables (**CMV_{12}**) is calculated as: $r_1 \cdot m_1 \cdot m_2 \cdot r_2$
- The method effect **m_i** can be calculated as: $\sqrt{(1 - v_i^2)}$
- The quality coefficients **q_i** can be calculated as: $r_i \cdot v_i$

The reliability and validity coefficients **r_i** and **v_i** can be obtained from:

SQP 2.0
Survey Quality Predictor

OUTLINE

Theory

Applicability using Stata

Benefits and possibilities

SPAIN'S CASE ESS ROUND 6

- **Regression model:**

$$Satdem = \alpha + \beta_L\, Lrplace + \beta_F\, Free + \beta_C\, Critic + \beta_E\, Equal + \beta_I\, Income + \zeta_S$$

- **Model variables:**
  - **Satdem:** Satisfaction with the democracy in Spain
  - **LRplace:** Self-placement on the left-right political scale
  - **Free:** Belief of freedom and fairness of elections in Spain
  - **Critic:** Belief of opposition parties' freedom to criticize the Spanish government
  - **Equal**: Belief that courts treat everyone the same
  - **Income:** Household income

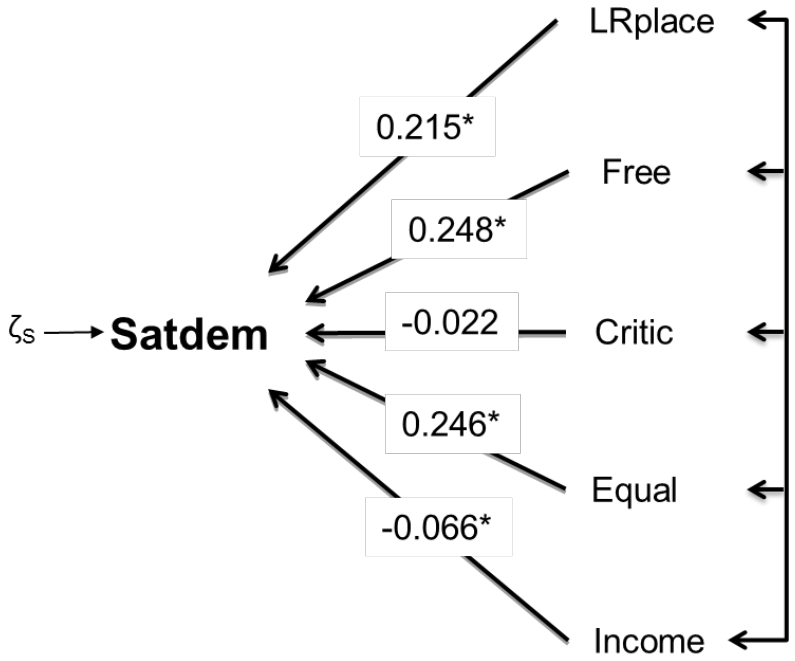
ANALYSIS WITHOUT CORRECTION FOR M.E.

- **We can analyse our model based on the correlation matrix using…**

```
ssd init satdem free critic equal lrplace income   /*variables*/
ssd set observations 1403                          /*observations*/
```

```
*Correlation matrix input
#delimit ;
ssd set correlations
1.000\
.3206 1.000\
.1173 .3429 1.000\
.3498 .2687 .1666 1.000\
.2873 .1083 .0809 .1954 1.000\
-.0275 .1392 .0560 .0164 .0072 1.000 ;
#delimit cr
```

```
*Regression model
sem (satdem <- free critic equal lrplace income), standardized
estat eqgof
```



- R^2: Only **22.6%** of the variance is explained

STEP 1: GET QUALITY INFORMATION

- We coded the characteristics of the 6 questions in our model using the **SQP 2 coding process**.
- The quality information is obtained:

	r	v	q	r^2	v^2	q^2	m
Satdem	0.895	0.956	0.856	0.801	0.914	0.733	0.293
Free	0.874	0.892	0.779	0.764	0.796	0.607	0.452
Critic	0.876	0.895	0.783	0.767	0.801	0.613	0.446
Equal	0.875	0.897	0.784	0.766	0.805	0.615	0.442
LRplace	0.858	0.940	0.807	0.736	0.884	0.651	0.341
Income	0.856	0.918	0.785	0.733	0.843	0.616	0.397

- Where method effect m_i is calculated as: $\sqrt{(1-v^2)}$

STEP 2: CORRECTION OF CORR MATRIX

- Observed correlation matrix without correction:

	Satdem	*Free*	*Critic*	*Equal*	*LRplace*	*Income*
Satdem	1.000					
Free	0.321	1.000				
Critic	0.117	0.343	1.000			
Equal	0.350	0.269	0.167	1.000		
Lrplace	0.287	0.108	0.081	0.195	1.000	
Inc	-0.028	0.139	0.056	0.016	0.007	1.000

$$\rho(f_1, f_2) = \frac{[r(y_1, y_2) - CMV_{12}]}{q_1 q_2}$$

- New correlation matrix corrected for measurement errors

	Satdem	*Free*	*Critic*	*Equal*	*LRplace*	*Income*
Satdem	1.000					
Free	0.481	1.000				
Critic	0.175	0.309	1.000			
Equal	0.521	0.190	0.025	1.000		
Lrplace	0.305	0.172	0.128	0.309	1.000	
Inc	-0.041	0.228	0.091	0.027	0.011	1.000

ESS EduNet

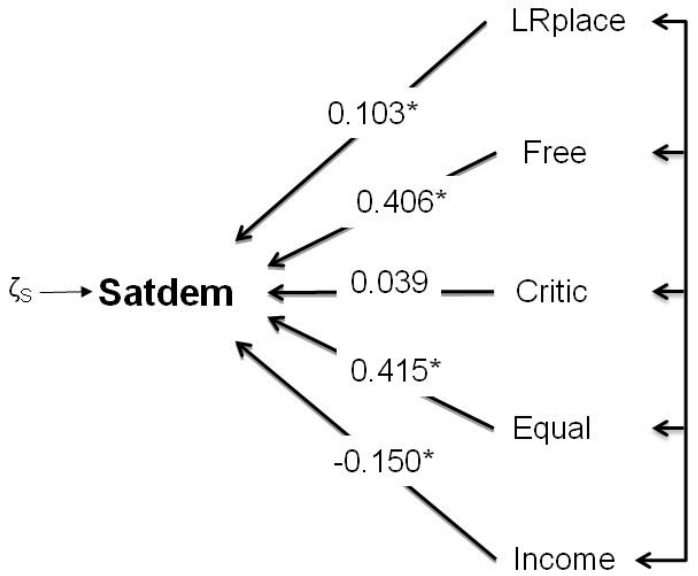
ANALYSIS WITH CORRECTION FOR M.E.

- **Analysing the new correlation matrix corrected for measurement errors using…**


```
ssd init satdem free critic equal lrplace income  /*variables*/
ssd set observations 1403                         /*observations*/
```

```
*Correlation matrix input
#delimit ;
ssd set correlations
1.00\
.481   1.00\
.175   .309   1.00\
.521   .190   0.025    1.00\
.305   .172   0.128   .309  1.00\
-.041  .228   0.091   .027   0.011   1.00 ;
#delimit cr
```

```
*Regression model
sem (satdem <- free critic equal lrplace income), standardized
estat eqgof
```



- R^2: Now **45.6%** of the variance is explained

COMPARING THE RESULTS WITH AND WITHOUT M.E.

	Results without corrections		Results with corrections	
Effects	**Coeff**	**E.Var**	**Coeff**	**E.Var**
Satdem <-		0.773		0.544
Free	0.248**		0.406** +0.158	
Critic	-0.022		0.039 +0.061	
Equal	0.246**		0.415** +0.169	
Lrplace	0.215**		0.103** -0.112	
Income	-0.066**		-0.150** +0.084	
R^2	**0.226 (22.6%)**		**0.456 (45.6%)**	

** if α<1% and * if 1%<α<5%

OUTLINE

Theory

Applicability using Stata

Benefits and possibilities

Benefits and possibilities

- **Benefits:**
- Your results will be better
- The R^2 of your model will increase.
- You don’t need to perform an experiment to test the quality of your measures.
- SQP is available online for free.
- Comparability across countries

- **Possibilties with Stata:**
- SEM is simple in Stata when the correlation or the covariance matrix is used.
- The covariance matrix can also be corrected for M.E. to obtain the unstandardized results.
- Different models that can be applied in Stata are illustrated in the Edunet module.

THANK YOU FOR YOUR ATTENTION!

Further information in:

"A simple procedure to correct for measurement errors in survey research"

Written by: Anna DeCastellarnau and Willem Saris

http://essedunet.nsd.uib.no/cms/topics/measurement/

anna.decastellarnau@upf.edu