Multiple imputation with Supervised Principal Component Regression

In one sentence

Using **supervised principal component regression** as a univariate **imputation** model in **MICE** is a great way to solve the **many-variables** imputation problem.

Large data with missing values (-)

	x_1	x_2	x_3	x_4	...	w_{141}	w_{142}	w_{143}	w_{144}	...	$z_{(p-3)}$	$z_{(p-2)}$	$z_{(p-1)}$	z_p
Esther	2	-	5	7		10	8	9	7		5	7	10	2
Anton	-	-	7	8		9	10	4	10		6	10	5	4
Leonie	-	6	-	4		3	7	8	3		3	6	10	9
Joran	6	3	1	-		10	6	4	1		2	1	6	10
...														
Mihai	-	9	-	9		6	9	10	3		7	3	7	4

Expert imputation model specification

- Remove constants and collinear variables.
- Evaluate connection between variables in the data.
- Apply a correlation-thresholding strategy.
- Extra: use total scores for item scales
- Extra: use single measurement in longitudinal data

VS

Automatic imputation model specification

- MICE with Principal component regression
- MICE with Association-threshold Supervised Principal Component regression
- MICE with Principal Covariates regression
- MICE with Partial least square

Percent relative bias

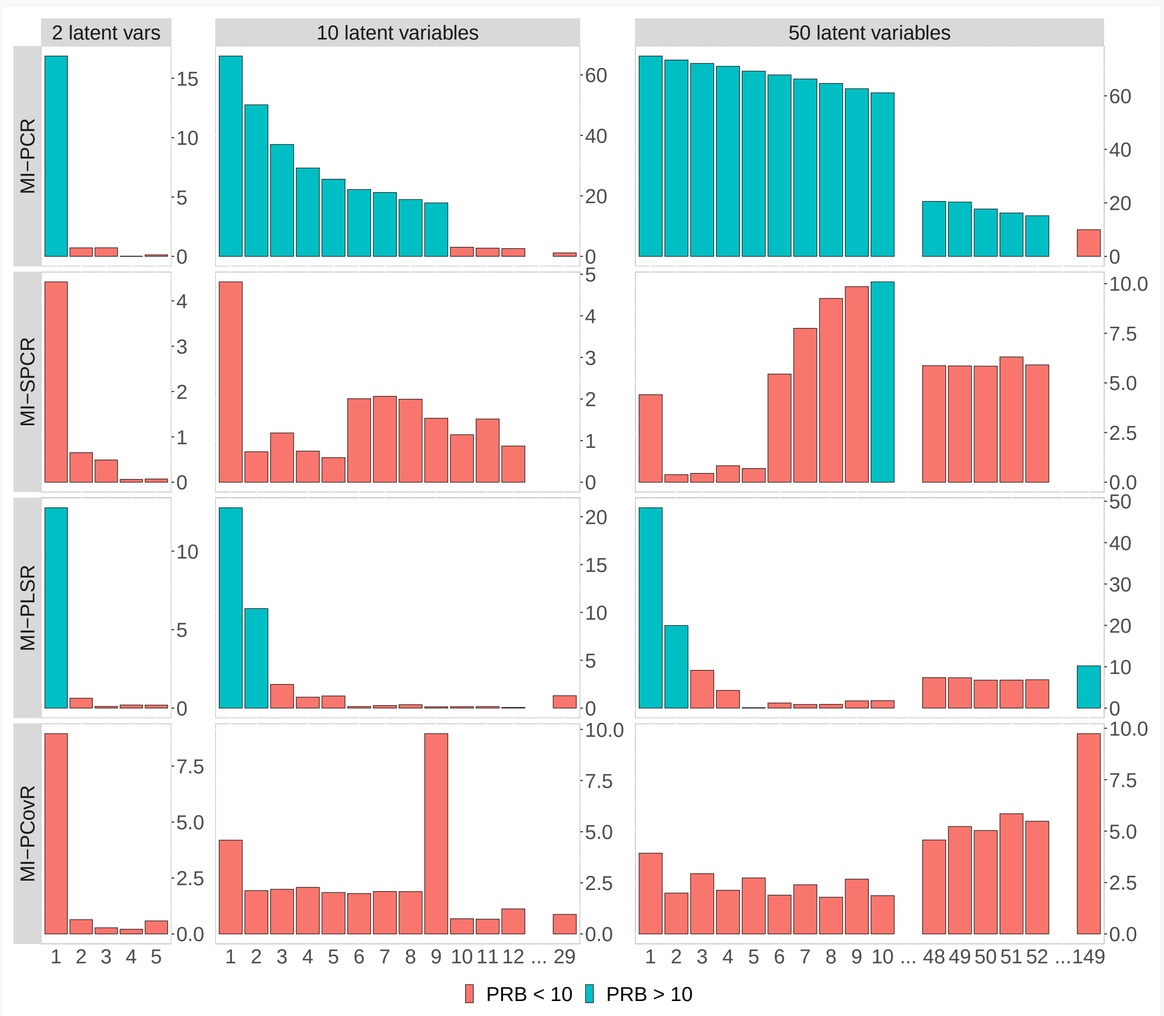

Figure: The percent relative bias (Y-axis) for the four PCR-based imputation methods (grid rows) is reported as a function of the number of components used (X-axis).

Confidence interval coverage

Figure: The confidence interval coverage for the four PCR-based imputation methods (grid rows) is reported as a function of the number of components used (X-axis).

Project summary and code

Play with the Shiny app

More research like this

Edoardo Costantini

e.costantini@tilburgunviersity.com

Tilburg University