

# Multiple imputation with Supervised Principal Component Regression

Using **supervised principal component regression** as an 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	1	-	6	2		4	1	1	2		5	2	1	2
Anton	-	-	9	5		8	8	1	10		4	3	5	9
Leonie	-	6	-	5		9	6	1	9		3	1	3	5
Joran	8	1	1	-		2	1	4	6		9	9	8	3
...														
Mihai	-	4	-	3		8	3	10	4		3	7	5	5

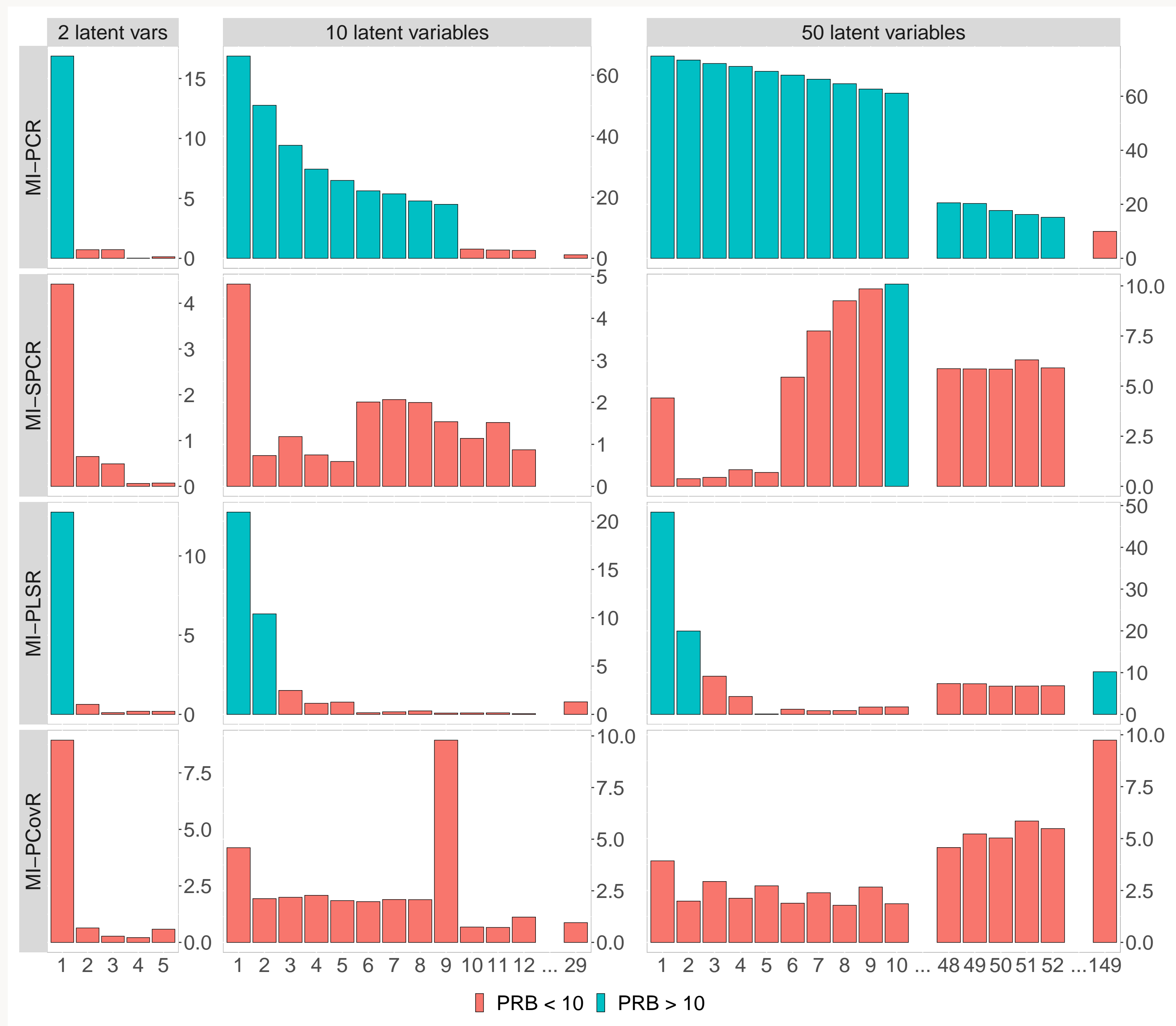
### Expert imputation model specification

- 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

### 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

