The impact of innovation on Healthcare costs: A multiple imputation approach

Erika Laranjeira[(1,2)] & Filipe Grilo[(1,2)]

(1) - FEP & (2) - CEF.UP, University of Porto, Portugal

2^{nd} Portuguese Stata User Group Meeting
Olisipo
September 7^{th}

Outline

Outline

Technological Innovation in Healthcare

Main source of development

Increases quality, safety, timeliness and efficiency of healthcare services

There is not a consensus concerning the potential benefits and costs savings

- New technologies are more expensive
- Their introduction can even increase the type and the number of treated patients

Technological Innovation in Healthcare

Newhouse (1992), for example, tried to measure the role of technological innovation in health expenditure growth

- Residual approach

This paper builds a technological index and tries to measure technological innovation directly

- This index requires technological health data from the 1980s
  - Problem:Missing Data

Multiple Imputation

Existence of missing values

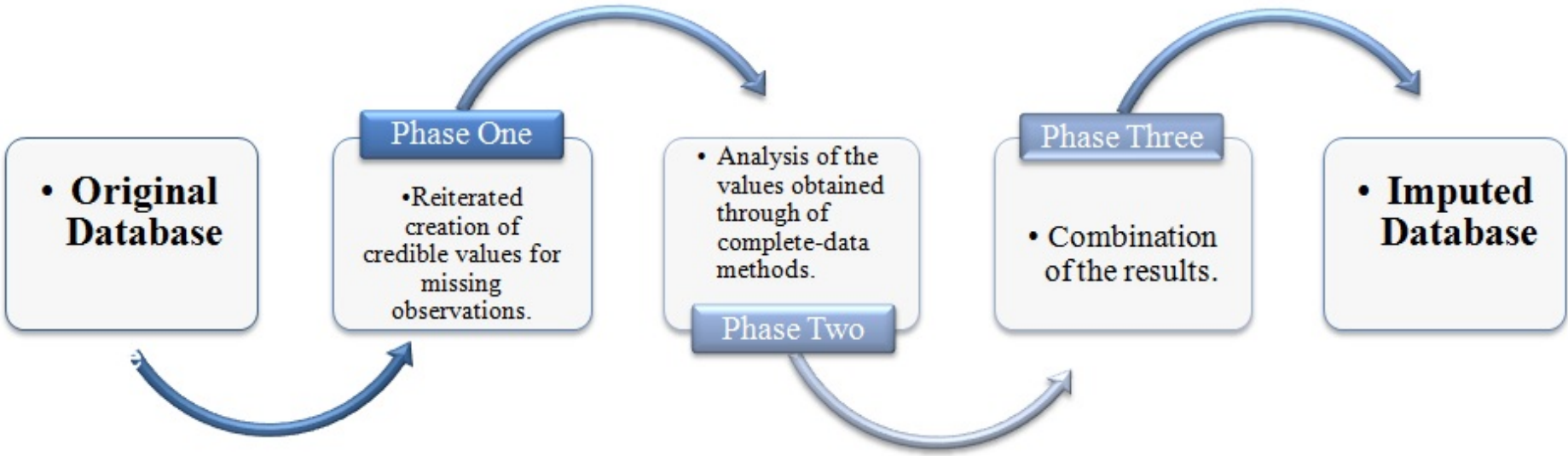
- Recurring problem in any real investigation
- Can compromise results
  - Current software assume complete database
    - Exclude, from the analysis, observations with missing values (Listwise Deletion)

The solution: Multiple Imputation

- It sees the missing values as an integrate part of the database and iteratively imputes them with values
- It creates $N \geq 2$ new databases

Outline

Multiple Imputation



mi set *style*
mi xtset *panelvar timevar* [, *tsoptions*]
mi register [imputed ; passive ; regular] *varlist*
mi impute mvn ivars [= *indepvars*]

Outline

The Technological Index

In order to measure healthcare technology, this paper had to create a technological index

- Quasi inexistence of quantitative data able to translate health technological level
- Main reference: TAI from the UN

16 variables

- Machines
- Procedures

The weights were determined by factor analysis

Skip to Pre-Estimation

Factor Analysis

Factor analysis groups together indicators that are collinear to form a composite indicator capable of capturing as much of common information of those indicators as possible
Process:

- First step: Through Principal-component factor, a matrix of factor loadings is created
  - factor *varlist* [*if*] [*in*] [*weight*] [, *method options*]
- Second step:Rotate the matrix of factor loadings
  - rotate
- Last step: Construction of the weights from the matrix of factor loadings after rotation and squaring it

Outline

Pre-Estimation: OLS vs. Robust

There was a previous suspicion of outliers and heterocedasticity within the data

If confirmed, Robust regression is more appropriate to these cases

To our knowledge, in STATA there is not a "Hausman test" in order to choose Robust over OLS

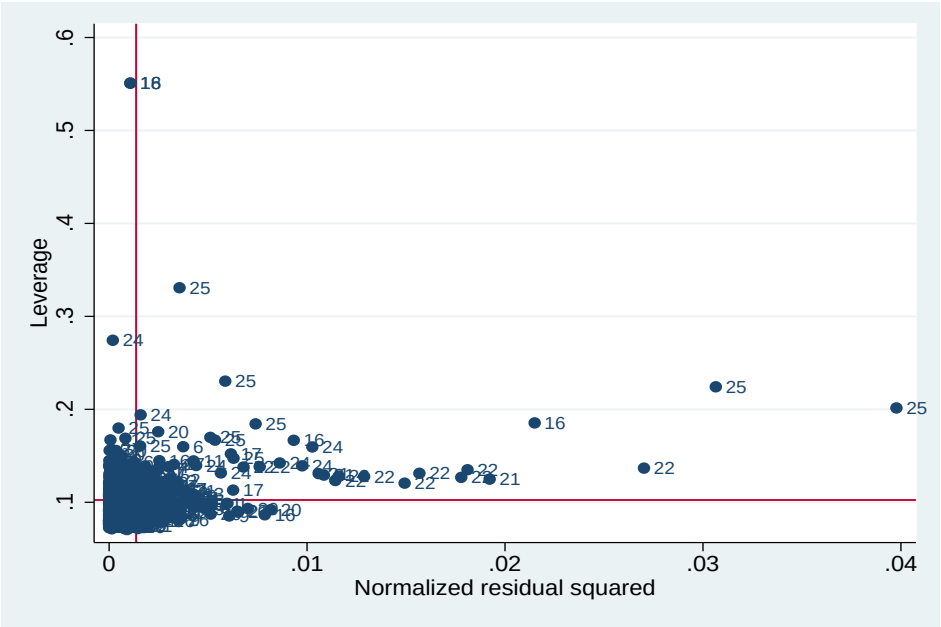
⏩ Skip to Est. Results

A simple algorithm

Simple algorithm:

http://www.ats.ucla.edu/stat/stata/dae/rreg.htm

- First step: OLS post-estimation diagnosis, focused on outliers
  - lvr2plot



A simple algorithm

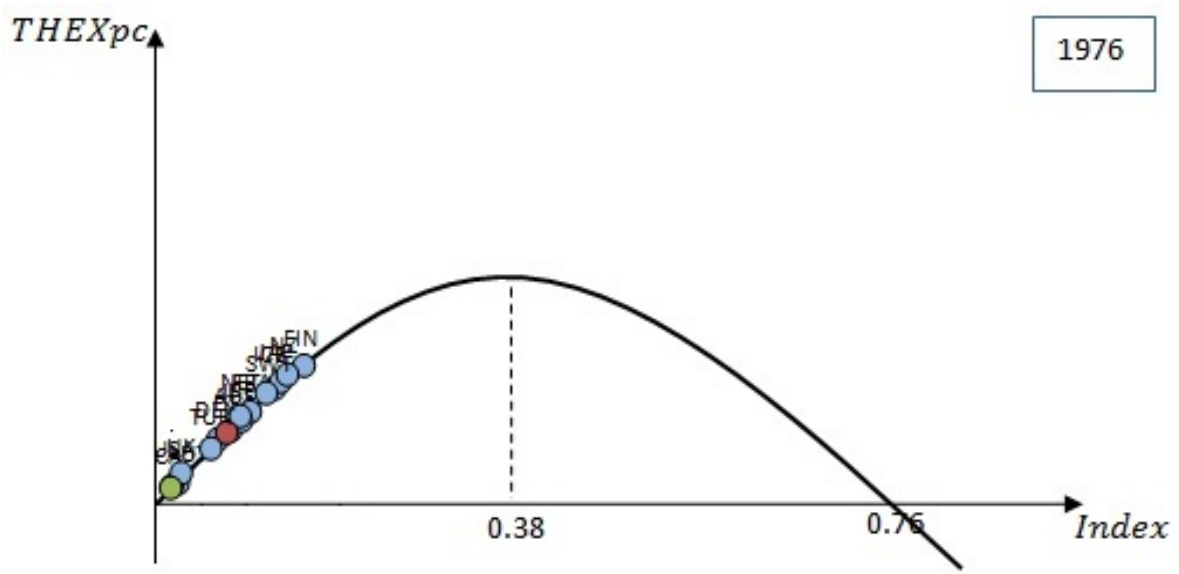
- Second step: Analyze the observations' weights attributed by Robust Regression
  - The more cases in the robust regression that have a weight close to one, the closer the results of the OLS and robust regressions
    - rreg *depvar* [*indepvars*], gen(weight)
    - list weight

Outline

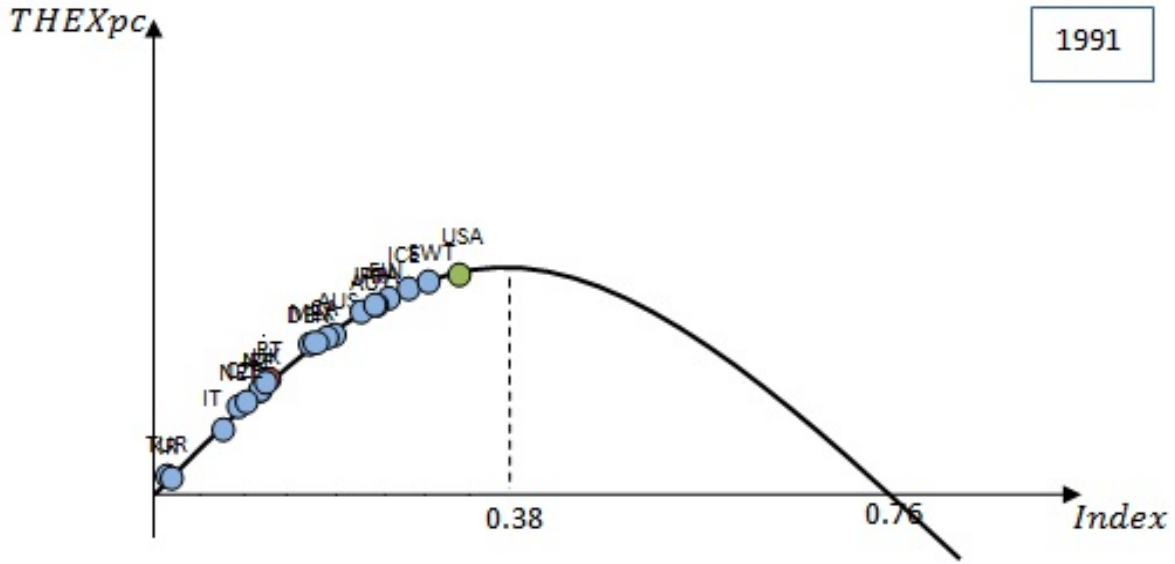
Estimation Results

Variables/Models	1	2	3
GDP	0.822^a	0.816^a	0.811^a
POP14	0.134^a	0.1313^a	0.127^a
EMV	0.016^b	0.0171^b	0.170^a
UP	-0.256^b	-0.236^b	-0.233^b
ALCOOL	0.05^a	0.053^a	0.051^a
SMOK	0.001	0.012	0.011
OUTP	-0.051^a	-0.053^a	-0.052^a
IM	-0.096^a	-0.098^a	-0.096^a
LEX		0.164	0.185
PHARM	0.015^c	0.014^c	0.0143^c
GOV	0.647^a	0.6415^a	0.642^a
GOOD			0.012
EMPLY	-0.012^b	-0.012^b	-0.014^b
INDEX	0.405^a	0.418^a	0.427^a
INDEXSQ	-0.534^a	-0.518^a	-0.522^a
Constant	-0.544	-1.28	-1.4
R^2	?	?	?
F-statistic	1973.88^a	1877.80^a	1746.26^a

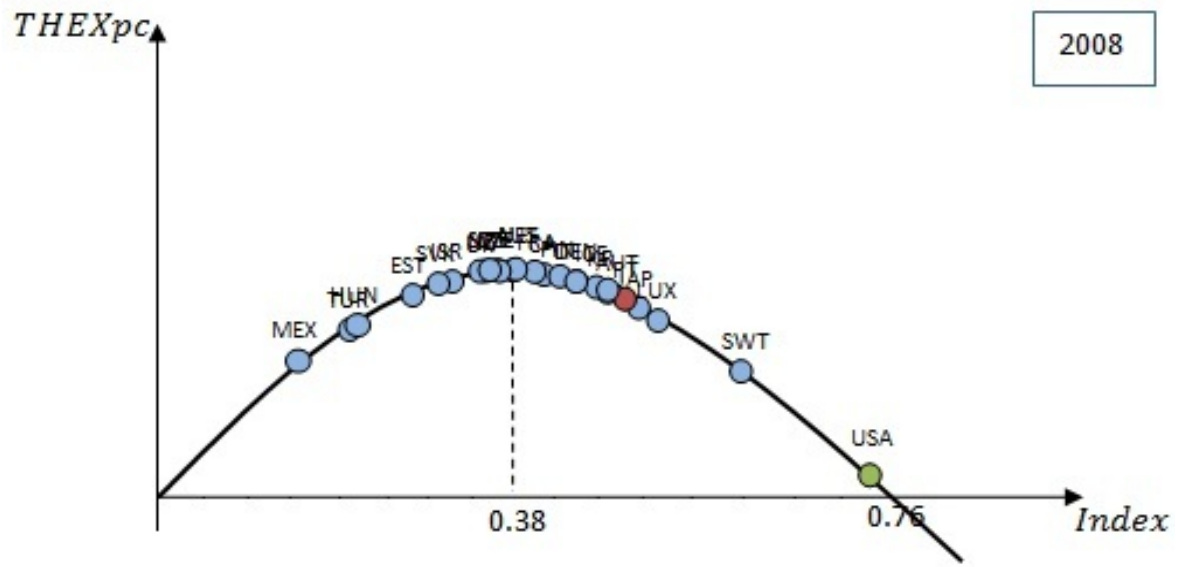
Technological Index



Technological Index



Technological Index



Outline

Problems

Using Multiple Imputation significantly limits the available tools to analyze results, since it is a recent package (available from STATA 11)
The use of robust regression exponentially aggravated this problem
Serious problems during this research:

- Factorial analysis
- Outlier detection
- Hausman Test (robust regression)
- R^2

Problems (and suggestions?)

Factor Analysis and Outlier Detection are not available after MI

- Solution: imputing 150 (or more) and taking the average
  - Working with asymptotically imputed variables
  - Underestimate the uncertainty of parameter estimation in the missing data case

Hausman test for robust regression

- For our knowledge, a code has not yet been developed

R^2

The R^2 case

- With *rreg* command, the R^2 given by STATA doesn't correspond to the true value
  - *rregfit* command is used instead
- With multiple imputation (even with just a single imputation, through the *mi xeq* command), the *rregfit* can't access to some particular data (?) and doesn't work

	150 Imp.	MI
e(R^2)	.9984	.9984
rregfit	.8167	...

The End

Thank you for your attention!