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\(^{th}\)
1 Motivation
   - Technological Innovation in Healthcare
   - Multiple Imputation

2 "Filling" the Database: Multiple Imputation

3 The Technological Index: Factor Analysis
   - The Technological Index
   - Factor Analysis

4 Pre-Estimation: OLS vs. Robust

5 Estimation Results
   - Estimation Results
   - Technological Index

6 Problems (and suggestions?)
Outline

1 Motivation
   - Technological Innovation in Healthcare
   - Multiple Imputation

2 “Filling” the Database: Multiple Imputation

3 The Technological Index: Factor Analysis
   - The Technological Index
   - Factor Analysis

4 Pre-Estimation: OLS vs. Robust

5 Estimation Results
   - Estimation Results
   - Technological Index

6 Problems (and suggestions?)
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
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

1 Motivation
   - Technological Innovation in Healthcare
   - Multiple Imputation

2 "Filling" the Database: Multiple Imputation

3 The Technological Index: Factor Analysis
   - The Technological Index
   - Factor Analysis

4 Pre-Estimation: OLS vs. Robust

5 Estimation Results
   - Estimation Results
   - Technological Index

6 Problems (and suggestions?)
Multiple Imputation

**Motivation**
"Filling" the Database: Multiple Imputation

The Technological Index: Factor Analysis
Pre-Estimation: OLS vs. Robust
Estimation Results
Problems (and suggestions?)

**Example Code**

```stata
mi set style
mi xtset panelvar timevar [, ts_options]
mi register [imputed ; passive ; regular] varlist
mi impute mvn ivars [= indepvars]
```
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.
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
There was a previous suspicion of outliers and heteroscedasticity 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
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

1. 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

1. Motivation
   - Technological Innovation in Healthcare
   - Multiple Imputation

2. "Filling" the Database: Multiple Imputation

3. The Technological Index: Factor Analysis
   - The Technological Index
   - Factor Analysis

4. Pre-Estimation: OLS vs. Robust

5. Estimation Results
   - Estimation Results
   - Technological Index

6. Problems (and suggestions?)
### Estimation Results

#### Variables/Models

<table>
<thead>
<tr>
<th>Variables/Models</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>0.822&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.816&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.811&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>POP14</td>
<td>0.134&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.1313&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.127&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>EMV</td>
<td>0.016&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.0171&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.170&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>UP</td>
<td>-0.256&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.236&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.233&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>ALCOOL</td>
<td>0.05&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.053&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.051&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>SMOK</td>
<td>0.001</td>
<td>0.012</td>
<td>0.011</td>
</tr>
<tr>
<td>OUTP</td>
<td>-0.051&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.053&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.052&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>IM</td>
<td>-0.096&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.098&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.096&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>LEX</td>
<td></td>
<td>0.164</td>
<td>0.185</td>
</tr>
<tr>
<td>PHARM</td>
<td>0.015&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.014&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.0143&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>GOV</td>
<td>0.647&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.6415&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.642&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>GOOD</td>
<td></td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>EMPLOY</td>
<td>-0.012&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.012&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.014&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>INDEX</td>
<td>0.405&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.418&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.427&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>INDEXSQ</td>
<td>-0.534&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.518&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.522&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.544</td>
<td>-1.28</td>
<td>-1.4</td>
</tr>
<tr>
<td>$R^2$</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>F-statistic</td>
<td>1973.88&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1877.80&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1746.26&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
Technological Index

THEXpc

Index

1976

0.38

0.76
Technological Index
Outline

1. Motivation
   - Technological Innovation in Healthcare
   - Multiple Imputation
2. "Filling" the Database: Multiple Imputation
3. The Technological Index: Factor Analysis
   - The Technological Index
   - Factor Analysis
4. Pre-Estimation: OLS vs. Robust
5. Estimation Results
   - Estimation Results
   - Technological Index
6. Problems (and suggestions?)
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
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

<table>
<thead>
<tr>
<th></th>
<th>150 Imp.</th>
<th>MI</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>e(R^2)</code></td>
<td>.9984</td>
<td>.9984</td>
</tr>
<tr>
<td><code>rregfit</code></td>
<td>.8167</td>
<td>...</td>
</tr>
</tbody>
</table>
Thank you for your attention!