



appraiseR: an R Package for Real Estate Appraisals

Luiz Fernando Palin Droubi
SPU

Willian Zonato
SPU

Norberto Hoccheim
UFSC

Abstract

The motivation behind **appraiseR** package comes from the need for better Real Estate Appraisals (REA) in Brazil, but not only. Brazilian standardization in this area is relative modern when compared to other countries, since it determines the use of inference methods whenever its possible to do so. On the other hand, the software industry in Brazil is not well advanced in this area. There are several different commercial softwares for REA, but none of them is able to do basic tests like Breusch-Pagan test for heteroskedasticity. Another problems comes from the softwares interfaces and the way people learn to use them, without any good statistical strategy. We pretend that with a free software package (and good interfaces built with shiny) we can improve teaching people to do better statistics.

Keywords: real estate, appraisals, R, shiny.

1. The appraiseR package

There is almost nothing in **appraiseR** built specially to Real Estate Appraisals. Code used to do inference for REA can be also be used in any other areas. Specially the shinyapps.

2. Statistical Strategy

Faraway (?, p.57) has a good approach to what can be a good statistical strategy. With simple tools widely available in R for exploratory analysis, diagnostic, transformation and variable selection one can almost ever find a good model. But, according to Faraway, its easy even to a well trained analyst to be confused when and where to use it, which can lead to not so good models and maybe worse predictions.

According to Faraway, there is no proper tool utilization order. An analysis may begin by the variable transformations, other from residual analysis of a first linear model, and so on.

Another way of starting an statistical analysis is with forward or backward variable selection. But Faraway recommends that the following order to be used:

Diagnostic→Transformation→Variable Selection→New diagnostic

It is not clear what Faraway mean by diagnostic, but we think is very important to emphasize that good exploratory can lead to good models. Faraway also says that is better to have in mind what objective has the statistical analysis: in REA we are always interested in good predictions.

3. The reality of REA

Problems with reliable data are always present when dealing with REA. Official transactions data are not always reliable, specially in Brazil since tax evasion is a constant, so the transactions prices are biased.

Advertising prices are fine, but it is clearly not the real value of the building, since there is always negotiation that leads to discounts.

We also think that dealing only with offers prices may also leads to Real Estate bubbles. In Brazil it did not happen any Minsky moment in Real Estate assets or something similar to USA. We think this is due to the fact that here there was not that sub-prime loans problem, but there is almost a consensus that there was a bubble but here and it is disinflating slowly, due to nominal rigidity prices.

4. Brazilian commercial softwares for REA

Brazilian commercial software have any power if compared to R. And teaching with commercial software is not the ideal procedure, first because we think code must always be available to critics, and teaching people with a commercial software is also not a good idea, since they might be interested in using another software.

Our initial strategy for solving this problem was to build an R package that could easily lead us to develop **shinyapps** with it. We consider that shiny reactive programming is very useful for teaching, and it is also difficult to convince people who are not familiar with programming to start using R, but shiny can be very attractive.

REApp was built with the intention to allow people who are already users of other commercial software to migrate to the use o R.

Briefly, it allows users to upload excel tables and uses the **bestfit appraiseR** function to build tables of several different models with different transformations for each variable in the model. In the sidebar user can control which variables will be used in the model, which transformations should be used to build the table, which model to test or use for predictions and the model outliers. Every chosen option affects reactive functions that updates the models automatically. Tests available for normality and homoscedasticity were chosen based in the standard, which includes Kolgomorov-Smirnov, Breusch-Pagan, and others. The majority of the tests comes from other packages, like **lmtest**, **normtest** and **nortest**, which is far beyond the tests available in the commercial software.

Special attention have been taken with exploratory analysis, which is a thing R can do better than any other software.

Functions to allow people to use box plots, histograms and scatter plots were made available by **REApp**, coming not only from **appraiseR** package but also from **mosaic** and others.

5. Benchmarking

6. Datasets available

There are a dozen of data sets extracted from previous teaching material, like `jungle`, `trindade` and others. Almost all of them are available with a specific purpose, i.e., for teach something different, like Spatial Regression techniques. Some of them are `SpatialPointsDataFrames`, like `centro_2015`, `centro_2013_2015`, `itacorubi_2015` and `trivelloni_2005`.

Affiliation:

Luiz Fernando Palin Droubi

SPU

First line Second line

E-mail: luiz.droubi@planejamento.gov.br

URL: <http://droubi.me>

Willian Zonato

SPU

First line Second line

E-mail: willian.zonato@planejamento.gov.br

Norberto Hoccheim

UFSC

First line Second line

E-mail: hochheim@gmail.com