Manual – Linear Regression v 1.1 Add-in

Jakub Rybacki, jakub.rybacki@gmail.com

Warsaw, 2014

Table of contents

[Introduction 3](#_Toc406577048)

[Installation 3](#_Toc406577049)

[Keyborard Shortcuts 3](#_Toc406577050)

[Requirements 3](#_Toc406577051)

[Diagnostics 4](#_Toc406577052)

[Implying Exogenous Variable 5](#_Toc406577053)

[Outliers Detection 6](#_Toc406577054)

[Filtering Outliers 7](#_Toc406577055)

[VIF Statistics 8](#_Toc406577056)

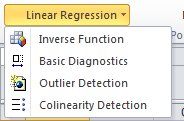
# Introduction

The aim of this document is to present functionality of Linear\_Regression Add-in, which extends function of Analysis Tool Pack Linear Regression, containing simple and user friendly diagnostic tools.

* Diagnostic test (RESET test, Breusch-Pagan test, Durbin-Watson test, Jarque-Bera test).
* Outliers Detection tools (based on Cook’s distance and leverage based statistics)
* Interface for Implying exogenous variable, knowing value of endogenous parameter (based on a goal seek function)
* Collinearity detection based on VIF statistics.

# Installation

Installation of Linear\_Regression Add-in would create following menu in the Add-ins tab at the ribbon:



Also functionality of some keyboard combinations would be overridden. Details below:

## Keyborard Shortcuts

|  |  |
| --- | --- |
| Shortcut | Action |
| CTRL + SHIFT + F8 | OLS Diagnostics |
| ALT + SHIFT + F8 | Exogenous variable implication |
| CTRL + ALT + F8 | Outliers Detection |

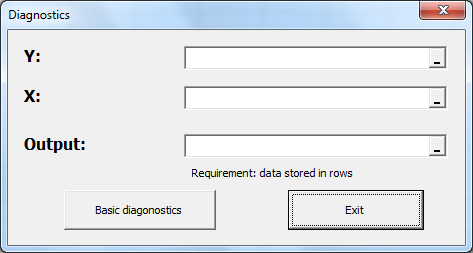
## Requirements

Time Series Add-in requires Solver installed. Official instruction how to install this add-in is available [here](https://support.office.com/en-nz/article/Load-the-Solver-Add-in-612926fc-d53b-46b4-872c-e24772f078ca). You may also find helpful following [video](https://www.youtube.com/watch?v=0z2BVsbITAs).

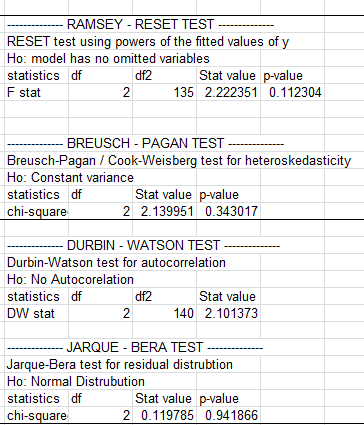
Please note also that every tool implemented in this add-in assumes that data is stored in rows (worksheet is column-oriented).

# Diagnostics

Selection of ***Basic Diagnostics*** option, or pressing **CTRL + SHIFT + F8** would trigger following interface:



Description of input field is quite intuitive – in *y* field we select full scope of endogenous variable, *x* field scope of exogenous variables, in *Output* field single cell referring to a place where output should be printed (output must be printed in the same worksheet were data is stored):

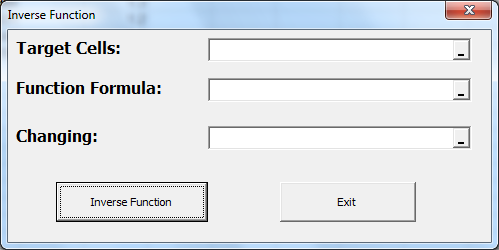


All stats were tested and are in line with Eviews results expects Jarque-Beta test. In case of JB test skewness and kurtosis yield similar results, however test formula is different. Linear Regression uses most common formula presented below:

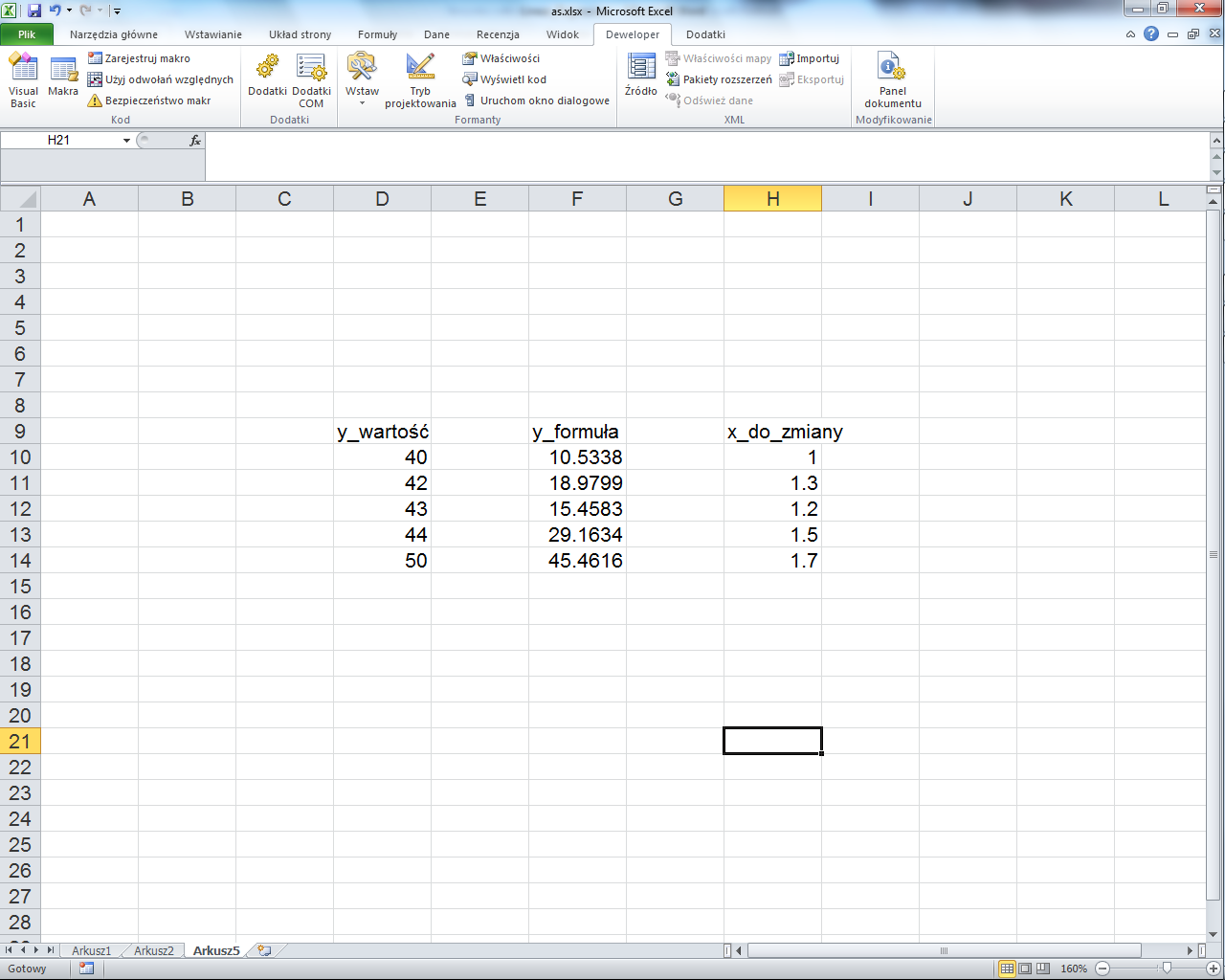

    \mathit{JB} = \frac{n}{6} \left( S^2 + \frac14 (K-3)^2 \right)
  

# Implying Exogenous Variable

Selection of ***Inverse Function*** option, or pressing **ALT + SHIFT + F8** would trigger following interface:



In field *Target Cells* scope of target values for endogenous variables shall be selected (this cells has to be pasted with *paste values* option). Field *Function formula* would contain similar scope by size, however in case of this cells formulas must be stored. Finally in *Changing* field scope of exogenous variables, which would be recalculated. All can be summarized by picture below:



Formula (in case of 10th row): =1.3\*POWER(H10;3)\*EXP(H10)+5\*H10+2

*(field function formula)*

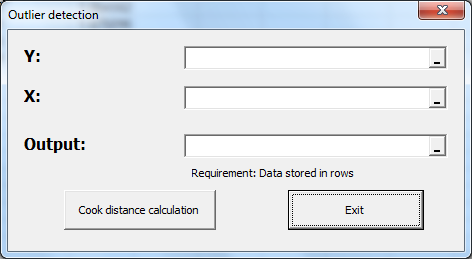
X – endogenous variables selected in field *changing*

Targeted values of Y pasted as values

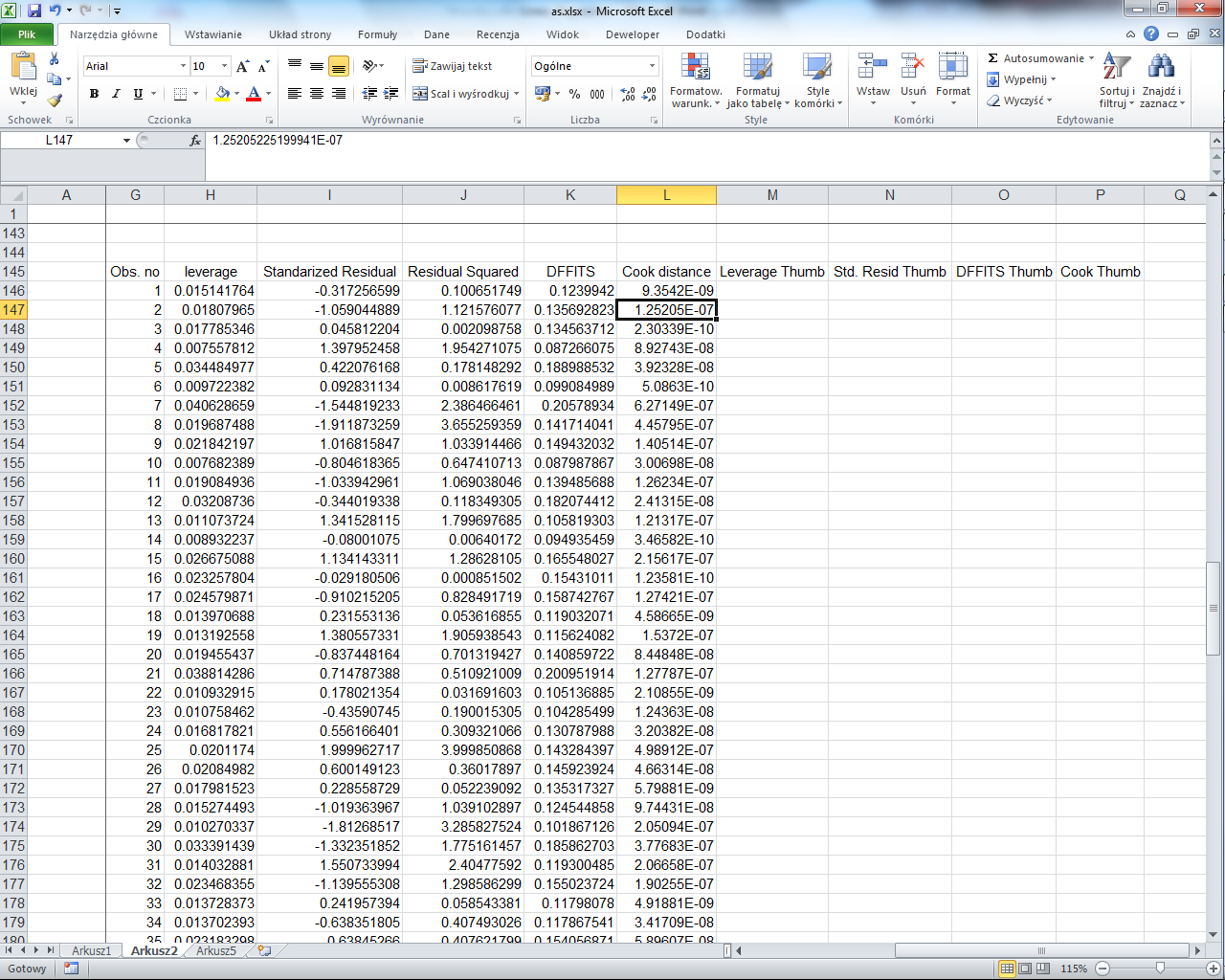
All X-es would be recalculated in such manner that column with formulas would have as similar values as it possible comparing to targeted Y-es. Macro assumes that whole data is stored in a single worksheet.

# Outliers Detection

Selection of ***Outlier Detection*** option, or pressing **CTRL + ALT + F8** would trigger following interface:



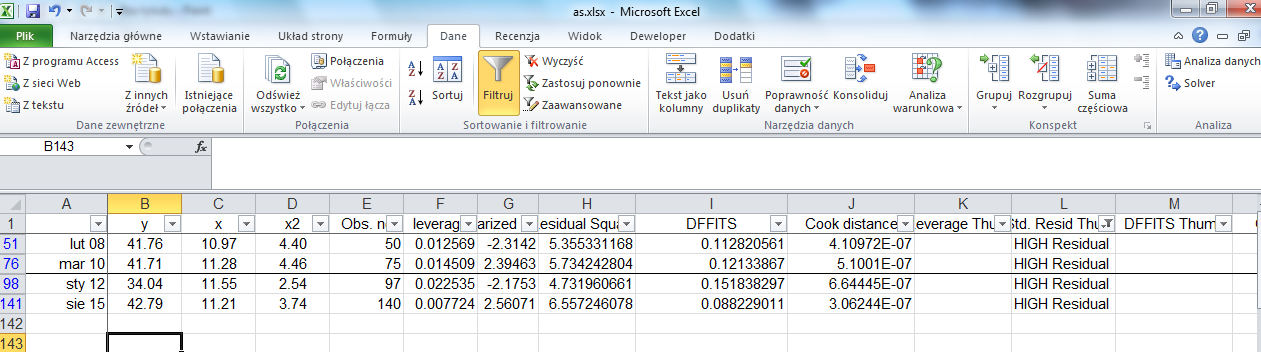
Input selection is similar like in case of Diagnostics – *y* contains endogenous variables, *x* exogenous variables, *Output* a single cell indicating, where output shall be printed (all should be stored in one worksheet). Output of macro based on randomly generated variables are presented below:



First 5 columns contains statistics explaining overall impact of observation on beta parameter. This are following: leverage statistic, standardized residual, its square, DFFITS statistics and Cook distance. Last five columns containing popular thumb rules allowing for quick outlier filtering.

## Filtering Outliers

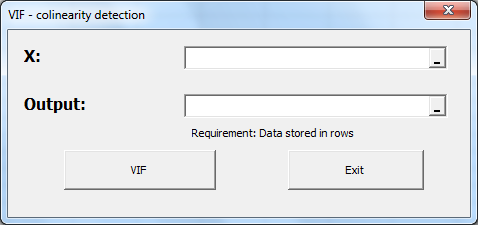
Before filtering outliers it is recommended to copy basic worksheet in order to prevent data loss (to do such operation right-click on a worksheet name and select move/copy worksheet). In the next step copy output of the outliers detection program adjacent to your x-es. Finally select option ***Data-> Filter*** and select heuristics based on which you would delete your outliers.



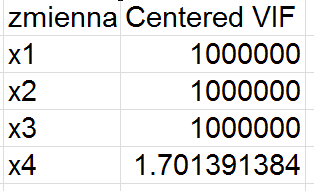
At the end delete rows with outliers.

# VIF Statistics

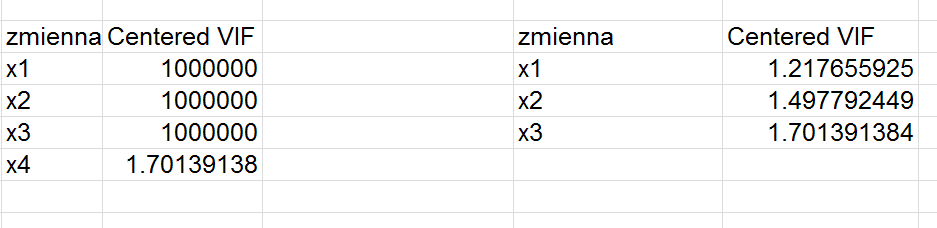
Selection of ***Colinearity Detection*** option would trigger following interface:



Field selection is quite intuitive –*x* field contains exogenous variables scope, *Output* single cell when output shall be printed (in the same worksheet). Let’s check output for randomly selected variables where x3=x2-x1:



Eliminating 3rd variable from output would yield acceptable results (note that variable formerly labeled as x4 became x3 due to elimination third variable from x-es):



Results were tested and were identical as Eviews outcomes.