Faculteit der Economische Wetenschappen en Bedrijfskunde

# SERIE   RESEARCH   MEMORANDA

## MetaGrowth 1 .O
## A  Computer  Program  for  Robustness  Analysis

Reinout   Heijungs
Henri L.F.  de Groot
Raymond J.C.M.  Florax

vrije   Universiteit    *amsterdam*

# MetaGrowth1.O
## A Computer Program for Robustness Analysis[*]

Reinout Heijungs, Henri L.F. de Groot, Raymond J.G.M. Florax

Department of Spatial Economics, Master-point
Free University
De Boelelaan 1105
108 1 HV Amsterdam, The Netherlands
E-mail: hgroot@feweb.vu.nl
Website: http://www.feweb.vu.nl/re/master-point

Metagrowth 1 .O — 28/08/0 1

# Contents

---

The program and example used in this manual
and
a pdf-version of this manual
can be downloaded from

http://www.feweb.vu.nl/re/mdw/henri or http://ww.feweb.vu.nl/re/master-point

---

# Introduction

## *Genera/ purpose*

*MetaGrowth* is a computer program facilitating analyses on the robustness of standard OLS regressions. It can be used to provide help to detect appropriate specifications, from a statistical perspective. As such, this is helpful in cases where theory does not provide sufficient guidance for model specification, or different (non-nested) operational definitions of variables are available.

For a data set with observations on a large number of variables an even larger number of possible model specifications exists. Some model specifications provide a good fit, some do not. Some variables have regression coeffcients significantly different from zero for almost all possible specifications, some do not. Some variables show significantly negative coeffcients for half of the model specifications but significantly positive coeffcients for the other half. Thus, in the context of uncertain or unknown model specifications, a detailed study of the robustness is an important extension to the traditional regression analysis. However, typically the robustness analysis is very time-consuming when performed with standard statistical software. It usually requires estimation of all possible specifications sequentially, one at a time. However, the summary statistics over all possible specifications cannot be obtained from standard statistical software. Consequently, keeping track of parameter estimates and standard errors manually makes for an extremely time-consuming task subject to considerable risk for errors.

*MetaGrowth* has been specifically designed to overcome these problems. *In* one step, *MetaGrowth* produces a table with regression results for *all* possible model specifications or a user-defined selection thereof. In addition, *MetaGrowth* automatically provides a range of robustness indicators.

## *Areas of Applica tion*

The need, desirability and benefits of robustness analysis extend to empirical problems in ecology, psychology and most other fields of science. One of the areas where the need for robustness analysis has been well established is the economic growth literature. This literature, expanding very rapidly over the past fifteen years, aims at empirically assessing the determinants of economic growth. Improved data availability and increased computer power eased the task of researchers to empirically test a wide variety of theories explaining economic growth. A seminal paper by Levine and Renelt (1992) seriously criticized this literature for its lack of robust insights. Since then, robustness analyses have been prominent features on the research agenda. For instance, Sala-i-Martin (1997) in turn criticized Levine and Renelt, maintaining that their criterion for robustness is too strict. The study by Florax et al. (2001) systematically assesses the robustness of various model specifications in the economic growth literature, and has been the impetus for the development of the *MetaGrowth* software. A more general treatment of robustness issues is given in Leamer (1983), Leamer and Leonard (1983), and McAleer et al. (1985).

## *Robustness indicators*

The Extreme Bounds Analysis developed by Leamer (1985) is the most frequently used method to assess the robustness of regression results. Extreme Bounds Analysis labels a relationship between a dependent variable y and an explanatory variable $x_i$ "robust," if the relationship between y and $x_i$ is of the same sign and statistically significant for *any* possible model specification containing $x_i$ and a set of explanatory variables $x_j$ ($j = 0, 1, \ldots, J$). Subsequent robustness analyses relaxed this requirement (see the discussion in Florax et al. 2001). Sala-i-Martin (1997), for example, introduced the criterion that the relationship should be significant in 95 percent of all cases. *MetaGrowth* makes it possible to experiment with different robustness criteria.

## Use of the program

*MetaGrowth* has been developed for academic purposes. Based on extensive testing of the software, the regression results of *MetaGrowth* are in accordance with regression results from other packages such as SPSS, STATA and EVIEWS. Furthermore, no computer stability problems occurred. However, we cannot guarantee that *MetaGrowth* operates reliably and smoothly under every circumstance and we cannot accept claims related to loss of data, incorrect analyses or any other problem.

# Use of MetaGrowth

## Starting MetaGrowth

The package is a small stand-alone executable program file, which runs under Microsoft Windows 95 and higher. There are no special memory or processor requirements, but a large RAM-size and a high clock speed obviously enhance the program's performance and possibilities.

The program can be started in the usual ways: (i) with Start | Run, (ii) with Windows Explorer, or (iii) by making and activating a desktop or toolbar shortcut to the program.

The program includes a demonstration data file, MetaGrowthDemo.txt. This data file contains a subset of the data provided and used by Sala-i-Martin (1997). The complete data set can be downloaded from his webpage: http://www.columbia.edu/%7Exs23/data/millions.htm. In this manual we use this data file for illustration purposes.

## Reading a data file

Upon starting the program, the tab sheet *Data* emerges. It contains space for the data to be analyzed, but on startup there is nothing to show. Click on *Open* to select a data file. If the file meets the format requirements, it is loaded and the data are shown on the screen.

The program expects tab-delimited text files. Columns indicate variables, rows indicate observations. Columns are separated by a tab, rows by a carriage return. The second column of the tab-delimited file should contain the dependent variable. Below is a small part of the data file MetaGrowth.txt that is used to illustrate the computations.

| # | Growth | gdpsh601 | lifee060 | p60 | safrica |
|---|--------|----------|----------|------|---------|
| 1 | 0.0136902 | 7.4389715 | 47.299999 | 0.46 | 0 |
| 2 | 0.0005685 | 6.7867169 | #N/A | 0.21 | 1 |
| 3 | -0.0065859 | 7.0192966 | 38.900001 | 0.27 | 1 |
| 4 | 0.0561947 | 6.2841344 | 45.700000 | 0.42 | 1 |
| 5 | 0.0042059 | 6.1527328 | 36.299999 | 0.08 | 1 |

The first column is entirely ignored by the program, even if it contains numerical information. The next columns (five in the example) contain the name of the column in the first row and numerical data in the next rows. Notice that dummy variables are allowed (see the variable *safrica* in the example). The value of *lifee060* of row (country) 2 is #N/A, indicating a missing value. All data items for which numerical parsing fails (such as 'NA', '???', '---', but also '12.3 (approx)' or '13-14') will be interpreted as a missing value. The table on the tab sheet *Data* shows the data entries as they have been interpreted, not as they appear in the file. All illegal expressions will show up as a blank entry. On completing the reading of a data file, the number of missing values, if any, is reported. The figure below shows the tab sheet *Data* after reading the demonstration data file.

General  Data  | Control regressions | Results of regressions | Analysis of regressions |

| nobs | growth | gdpsh6 | lifee060 | p60 | safrica | laam | infl | gpop | assass | war | democ | catholic | mining |
|------|--------|--------|----------|-----|---------|------|------|------|--------|-----|-------|----------|--------|
| 1 | 0.0137 | 7.44 | 47.3 | 0.46 | 0 | 0 | 8.06 | 0.0284 | 0 | 0 | 0.272 | 0 | 0.196 |
| 2 | 0.0005E6.79 | | | 0.21 | 1 | 0 | 9.46 | 0.0241 | | 1 | 0.12 | 0.38 | 0.268 |
| 3 | -0.0065! | 7.02 | 38.9 | 0.27 | 1 | 0 | 6.35 | 0.024 | | 0 | 0.247 | 0.07 | 0.009 |
| 4 | 10.0562 | 6.28 | 45.7 | 0.42 | 1 | 0 | 7.86 | 0.0331 | 0 | 0 | | 0.25 | 0.533 |
| 5 | 10.00421 | 6.15 | 36.3 | 0.08 | 1 | 0 | 4.93 | 0.0232 | 0 | 0 | 0.072 | 0.1 | 0.001 |
| 6 | -0.0026 | 6.38 | 41.8 | 0.18 | 1 | 0 | 5.8 | 0.0203 | 0 | 1 | 0.58 | 0.62 | 0.01 |
| 7 | | | | | | 0 | 7.12 | 0.0263 | 0 | 0 | 0.56 | 0.16 | 0.088 |
| 8 | 0.0122 0.0017 | 6 | 43.4 | 0.65 | 1 | 1 | 0 | 0.0206 | | | | | 0.006 |
| 9 | -0.0082 | 6.49 | 39.3 | 0.32 | 1 | 0 | 7.64 | 0.0226 | 0 | 0 | 0.342 | 0.25 | 0.03 |
| 10 | -0.0181 | 6.5 | 34.9 | 0.17 | 1 | 0 | 5.68 | 0.0202 | 0 | 1 | 0.476 | 0.16 | 0.002 |
| 11 | 0.00017 | 6.25 | | | 1 | 0 | | 0.0311 | | | | 0.14 | 0.079 |
| 12 | 0.0223 | 6.97 | 47.3 | 0.78 | 1 | 0 | 6 | 0.0275 | 0.188 | 0 | 0.458 | 0.25 | 0.168 |
| 13 | 0.0267 | 6.65 | 46.4 | 0.66 | 0 | 0 | 8.27 | 0.0238 | 0.006811 | | 0.387 | 0 | 0.034 |
| 14 | -0.0005! | 5.52 | 42.2 | 0.07 | 1 | 0 | 2.66 | 0.0228 | 0 | 1 | 0.125 | 0 | 0.001 |
| 15 | 0.0225 | 7.49 | 40.9 | 1 | | 1 | 0 | 9 | 0.0302 | 0 | 0 | 0.447 | 0.32 | 0.215 |
| 16 | 0.0095E | 6.2 | 32.3 | 0.12 | 1 | 0 | 8.15 | 0.03 | 0 | 0 | | 0.04 | 0 |

| 0 Open...e | n | results... | | | | Print... |

Due to differences in regional settings of the operating system, one may sometimes wish to load a data file with a decimal comma while the program expects a decimal point, or the other way around. To deal with this, the tab sheet *General* contains a field Decimal separator.

Data cannot be edited in *MetaGrowth*. For this purpose, one can use the DOS editor, or alternatively any general word processing or spreadsheet software (the latter also opens the possibility to put the output in a better format). Data can be spooled to the printer by pressing the Print button. The *Open results* button enables the user to load a previously saved file with regression results from an earlier analysis with *MetaGrowth*. It is further discussed in the section on regression results.

## Regression analyses

The tab sheet *Control regressions* allows the user to determine which model specifications the program should analyze.

First, it shows the name of the dependent variable, in this example the variable 'growth' (recall that the dependent variable must always appear in the second column of the data file). The upper panel of the screen also shows the number of rows of the data file (Observations) and the number of model specifications for which regressions will be estimated (Runs). Clearly, the number of 'Runs" will change as the choices of model specifications changes.

Dependent
growth

Sizes
Observations: 134      Runs: 41

Regressors
Regressors: 12

Switch limit
Upper limit: [5] ⟳
Lower limit: [3] ⟳

☐ Balanced design
☑ Standardize regressors

Include: 2    Include all    Switch: 6    Switch all    Exclude: 4    Exclude all

| Include | Switch | Exclude |
|---|---|---|
| ☑ gdpsh60l | ☐ gdpsh60l | ☐ gdpsh60l |
| ☐1 lifee060 | ☑ lifee060 | ☐ lifee060 |
| ☐ p60 | ☑ p60 | ☐ p60 |
| ☐ safrica | ☑ safrica | ☐ safrica |
| ☐ laam | ☐ laa | ☑ laam |
| | ☐ infl | ☐ infl |
| ☐ gpop | ☑ gpop | 3 gpop |
| ☐3 assass | ☐ assass | ☑ assass |
| ☐ war | ☐ war | ☑ war |
| ☐ democ | ☑ democ | ☐ democ |
| ☐ catholic | ☑ catholic | ☐ catholic |
| ☐ mining | ☐ mining | |

Results
☑ To table      ☐ To file: [ ]

The lower panel of the tab sheet allows the user to choose the independent variables (*Regressors*) of the regressions to be analyzed. There are three checklists, each containing the names of all regressors in the data set. The three lists are referred to as *Include,* Switch and *Exclude.* The checked variables in the *Include* panel will be included in all model specifications. The checked variables in the *Exclude* panel will be excluded in all model specifications. The checked variables in the *Switch* panel are those of primary interest for the regression analysis and will be included in some of the model specifications and excluded in other model specifications.

Let US consider a simple example with six regressors labeled A, B, C, D, E and F, respectively. Suppose we categorize A and B as included variables, C, D and E as switch variables, and F as an excluded variable. As a result, there will be eight possible model specifications: (A, B), (A, B, C). (A, B, D), (A, B, E), (A, B, C, D), (A, B, C, E), (A, B, D, E) and (A, B, C, D, E). In general, the number of possible model specifications $N_m$ is:

$$N_m = 2^{n_s},$$

where $n_s$ denotes the number of switch variables. Thus, if $n_s$ is large, the number of possible regressions will be quite extensive. This potentially creates computer time and memory constraints. It is therefore recommended to take advantage of the *Switch limit:* the upper limit represents the maximum number of switch variables included in the model specifications, whereas the lower limit represents the minimum number of switch variables included in the model specifications. Using the example above and setting the lower limit to 1 and the upper limit to 2, the set of model specification is restricted to (A, B, C), (A, B, D), (A, B, E), (A, B, C, D), (A, B, C, E) and (A, B, D, E). Models with less than three or more than four regressors will not be evaluated. In general, denoting the lower limit by $n_l$ and the upper limit by $n_u$, the number of model specifications reduces to:

$$N_{\mathrm{m}} = \sum_{i=n_l}^{n_u} \binom{n_s}{i},$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

is the binomial factor. Re-iterating the case of 17 switch variables, but setting the upper and lower limits to 6 and 0, respectively, reduces the number of model specifications to slightly more than 20,000, a reduction by a factor greater than 6. The use of switch limits is not only recommendable from an operating point of view; it is also sensible given that in econometric practice the number of independent variables will rarely exceed ten.

The tab sheet also contains a checkbox to *Standardize regressors*. If checked, all estimations are based using the standardized regressors. Standardization is recommended if the regressors vary substantially in magnitude, as large variations in magnitude typically yield poor performance of the matrix estimation algorithms and considerable errors in the estimates. This is the case if, for instance, one column indicates the length of the railway tracks (in km) and another column indicates the fraction of Nobel Prize winners of a population. The disadvantage of standardization is a longer computation time.

Typically, however, standardization yields equivalent results to the non-standardized version, except for the condition number. The condition number C is defined as:

$$C = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}},$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest eigenvalue of the variante-covariance matrix of the regressors. Alternatively, the condition number can be defined as $C = \lambda_{\max}/\lambda_{\min}$, where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest eigenvalue of the moment matrix of the regressors. The program will show a value of $10^{20}$ when $\lambda_{\min}$ is zero or has been rounded to zero, i.e., when C is infinite. Large condition numbers, say C $10^6$, indicate that the results of the computation are suspicious. In that case, it is advisable to redo the regressions with standardized regressors. For standardized regressors a condition number in excess of 20 shows that multicollinearity is a problem (see Belsley et al., 1980).

The option referring to *Balanced design* deals with missing values. The estimations exclude all observations for which at least one variable has missing data. This procedure may, however, lead to unequal sample sizes among the regressions if missing values of the regressors are recorded for different observations. Checking the *Balanced design* box ensures that all observations with missing values for any of the regressors are discarded in all model specifications.

The last entries of the panel allow users to choose where to store the results. There are two options. The first stores the results in a table that is accessible at the tab sheet *Results of regressions*, and can be saved at a later stage. The second option stores the results in a file. This option is recommended when the analysis involves a large number of regressions as it saves memory. Moreover, each new regression result is added directly, so that a program crash or power failure will not result in loss of regression results.

The tabsheet General has a field *Show run during progress*. Checking this field results in a counter indicating the number of models already computed. Computational speed may, however, be lowered.

## Inspecting the regression results

When selecting the tab sheets *Results of Regressions* or *Analysis of Regressions,* the calculations begin and a progress indicator appears. Clicking the *Cancel* button will immediately interrupt the calculation. All the results up to that point will be saved, but the program does not allow continuation at the point of interruption.

If the regression results are stored in a table, they can be accessed via the tab sheet *Results of regressions.* They can also be printed or saved to a data file (as tab-delimited text). If the results are stored in a file, they must be accessed by clicking the *Open results* button at the tab sheet *Data. In* each case, the table can be very large depending on the number of model specifications and/or regressors. The estimation of each model specification is reported in a separate row, and for each model specification, the columns show:

- number of observations (*nobs*);
- number of regressors *(nreg);*
- regression coeffcients for the constant and all regressors;
- t-values (test against zero) for all estimates;
- summary statistics:
  - 0 total sum of squares *(tss);*
  - o explained sum of squares (ess);
  - o residual sum of squares (*rss*);
  - o coefficient of determination ($r^2$);
  - 0 *F*-value (*f*);
  - o condition number of the matrix (*condnum*);
  - o log likelihood measure of goodness-of-fit (*ln likelihood).*

**MetaGrowth** — © Heijungs, De Groot, Florax — 8

The regression results may again be saved or printed. Printout of better quality is obtained by saving the output, and subsequently importing it into a spreadsheet.


## *Analyzing the regression results.*

Selecting the tab sheet *Analysis of regressions* yields a statistical analysis of all regression models. The analysis results are shown in a large table where the rows are defined by the constant and the regressors. This table may be printed or saved. The columns report the statistical properties for the parameter estimates including, for example, the number of models in which this particular regressor has been used, the mean value of the estimated coefficients of all these models, and the variance of the estimated coefficients. Table 1 below lists the statistics that are computed. In the tab sheet General, a field Full analysis of regressions controls whether all columns are reported or, alternatively, only a small number of the most important columns are shown. In Table 1 these are indicated with an asterisk.



| 𝒻 MetaGrowth 1.0 (demo version) | | | | | | _ □ ⊠ |
| --- | --- | --- | --- | --- | --- | --- |

General | Data | Control regressions | Results of regressions   Analysis of regressions |

| 'ariable | number of m | mean value | kft side of confidence inte | right side of confidence int | fraction of positive values | fraotion of significant |
| --- | --- | --- | --- | --- | --- | --- |
| :ONST | 64 | 10. 0333 | 0. 0171 | 0. 0496 | 0. 922 | 0. 547 |
| xdpsh60I | 64 | -0. 00656 | -0. 0103 | -000283 | 0. 125 | 0. 0625 |
| fee060 | 32 | 0. 00109 | 0. 000803 | 0. 00138 | 1 | 0. 938 |
| ι60 | 32 | 0. 0325 | 0.0246 | 0. 0403 | 1 | 1 |
| :africa | 32 | -0. 0154 | -0. 0192 | -0. 0117 | 0 | 0 |
| nfl | 64 | -2.32E-5 | -3.34E.5 | -1.x-5 | 0. 0625 | 0 |
| jpop | 32 | -0.204 | -0.414 | 0. 00603 | 0.281 | 0 |
| jemoc | 32 | -0. 00656 | -0. 0106 | -0. 00251 | 0. 188 | 0 |
| :atholic | 32 | -0. 0106 | -0. 0132 | -0. 00796 | 0 | 0 |

◄⌷                                                                                ►⌷

Export | Print

The items of the statistical analysis involves (two-sided) hypothesis testing where the default significance level is $\alpha = .05$ . The default significance level can be changed by accessing the tab sheet General, and altering the entry in the field Significance level. Changing the default significance level will automatically adjust some of the column headings that refer to confidence intervals. In addition, the user can also choose between tests based on an exact critical t-value referring to the pre-specified significance level, and tests with a constant t-value of $t = 2$. Using the exact t-value is recommended, but for reasons of compliance with Sala-i-Martin (1997), testing may also proceed with a t-value of two.

The table can be printed or saved (again, use a spreadsheet for better quality).

**Table 1:** list of statistics, their meaning, and computational details."

| Column name | Meaning | Equation |
|---|---|---|
| number of models* | Number of model specifications in which this variable has been included | $N$ |
| mean value* | mean value of the estimates of the regression coefficient for this variable | $\bar{b} = \dfrac{1}{N}\sum_{i=1}^{N} b_i$ |
| Variance | sample variance of the estimates | $s^2 = \dfrac{1}{N-1}\left(\sum_{i=1}^{N}(b_i)^2 - N\bar{b}_i\right)$ |
| standard deviation | sample standard deviation of the estimates | $s = \sqrt{s^2}$ |
| standard error of mean | sample standard error of the estimated mean | $se = \sqrt{\dfrac{s^2}{N^2}}$ |
| probability of mean | probability that the estimated mean differs from zero | $p = p\left(t = \dfrac{\bar{b}}{se}, df = \text{N-1}, \text{twosided}\right)$ |
| let? side of confidence interval* | left side of symmetric interval that encloses the estimated mean a) with probability 1-p$_{crit}$, b) corresponding to a t-value of 2 | $b_{ci-} = \bar{b} - t_{crit} \times se$, with $t_{crit} = t(p = p_{crit}, df = N-1, \text{twosided})$ or $t_{crit} = 2$ |
| right side of confidence interval* | right side of symmetric interval that encloses the estimated mean | $b_{ci+} = \bar{b} + t_{crit} \times se$ |
| left side of range | smallest estimate of the regression coefficient | $b_{min} = \min_{i=1}^{N} b_i$ |
| right side of range | largest estimate of the regression coefficient | $b_{max} = \max_{i=1}^{N} b_i$ |
| number of negative values | number of estimates with a negative regression coefftcient | $N_{neg} = \sum_{i=1}^{N} \Theta(-b_i)$ with $\Theta(x) = \begin{cases} 1 \text{ if } x > 0 \\ 0 \text{ otherwise} \end{cases}$ |
| fraction of negative values | fraction of estimates with a negative regression coefficient | $f_{neg} = \dfrac{N_{neg}}{N}$ |
| number of positive values | number of estimates with a positive regression coefftcient | $N_{pos} = \sum^{N} \Theta(b_i)$ |

| Column name | Meaning | Equation |
|---|---|---|
| fraction of positive values* | fraction of estimates with a positive regression coefficient | $f_{pos} = \dfrac{N_{pos}}{N}$ |
| left side of bounds-range | smallest (estimate minus two t-values) | $b_{br-} = \min_{i=1}^{N}\left(b_i - 2\dfrac{b_i}{t_i}\right)$ |
| right side of bounds-range | largest (estimate minus two t-values) | $b_{br+} = \min_{i=1}^{N}\left(b_i + 2\dfrac{b_i}{t_i}\right)$ |
| strong sign test (all equal sign) passed? (yes=1) | yes if all estimates have the same sign | 1 if $N_{pos} = N$ or $N_{neg} = N$ |
| weak sign test (95% equal sign) passed? (yes=1) | yes if 95% of estimates have the same sign | 1 if $N_{pos} \geq (1 - p_{crit})N$ or $N_{neg} \geq (1 - p_{uit})N$ |
| number of significant values | number of estimates that differ signiticantly from zero | $N_{sig} = \sum_{i=1}^{N} \Theta(abs(t) - t_{crit})$ |
| fraction of significant values* | fraction of significant estimates that differ from zero | $f_{sig} = \dfrac{N_{sig}}{N}$ |
| number of significant positive values | number of significant estimates that is larger than zero | $N_{possig} = \sum_{i=1}^{N} \Theta(t - t_{crit})$ |
| fraction of significant positive values* | fraction of significant estimates that is larger than zero | $f_{possig} = \dfrac{N_{possig}}{N}$ |
| number of significant negative values | number of significant estimates that is smaller than zero | $N_{negsig} = \sum_{i=1}^{N} \Theta(-t - t_{crit})$ |
| fraction of significant negative values | fraction of estimates that is significantly smaller than zero | $f_{negsig} = \dfrac{N_{negsig}}{N}$ |
| strong extreme bounds test (all significant and equal sign) passed? (yes=1) | yes if all estimates have the same sign and differ significantly from zero | 1 if $N_{possig} = N$ or $N_{negsig} = N$ |
| weak extreme bounds test (95% significant and equal sign) passed? (yes= 1) | yes if 95% of estimates have the same sign and differ significantly from zero | 1 if $N_{possig} \geq (1 - p_{crit})N$ or $N_{negsig} \geq (1 - p_{crit})N$ |

| Column name | Meaning | Equation |
|---|---|---|
| weighted mean value | mean value of the estimates weighted with the loglikelihood of the regression | $\overline{b}_w = \dfrac{1}{N}\dfrac{\sum\limits_{i=1}^{N} w_i b_i}{\sum\limits_{i=1}^{N} w_i}$ with $w_i$ the loglikelihood of the regression |
| weighted variance | weighted sample variance of the estimates | $s_w^{\,2} = \dfrac{\sum\limits_{i=1}^{N} w_i \left(b_i \big/ t_i\right)^2}{\sum\limits_{i=1}^{N} w_i}$ |
| weighted standard deviation | weighted sample standard deviation of the estimates | $s_w = \sqrt{s_w^{\,2}}$ |
| weighted standard error of mean | weighted sample standard error of the estimated mean | $se_w = \sqrt{\dfrac{s_w^{\,2}}{N^2}}$ |
| weighted probability of mean | probability that the estimated weighted mean differs from zero | $p_w = p\!\left(t = \dfrac{\overline{b}_w}{se_w}, df = N-1, \text{twosided}\right)$ |
| left side of weighted confidence interval | left side of symmetric interval that encloses the estimated weighted mean | $b_{wci-} = \overline{b}_w - t_{crit} \times se_w$ |
| right side of weighted confidence interval | right side of symmetric interval that encloses the estimated weighted mean | $b_{wci+} = \overline{b}_w + t_{crit} \times se_w$ |
| weighted extreme bounds test (95% significant and equal sign) passed? (yes= 1) | yes if the weighted confidence interval does not enclose zero | $1$ if $\left(\overline{b}_w - t_{crit} \times se_w\right)\times\left(\overline{b}_w + t_{crit} \times se_w\right) > 0$ |
| weighted CDF | cumulative density function; fraction that lies at the right side of zero | $CDF_w = 1 - p\!\left(z = \dfrac{\overline{b}_w}{s_w}, \text{onesided}\right)$ |

[a] An asterisk (*) indicates that the column appears in the summary analysis as well.

### Quitting MetaGrowth

The program may be stopped with the "X " in the title bar of the window, using ALT-F4, or with the Exit button at the tab sheet *General*.

# Technical details

### Implemen ta tion

MetaGrowth has been written in Delphi, a visual object-oriented version of Pascal. All computational routines are w-ritten in Pascal. Most algorithms have been taken from Press et al. (1989). The regression coefficients are obtained with the use of singular value decomposition, as this is the most robust implementation for a general regression problem.

### Performance

MetaGrowth has been extensively tested with a small and a large data file. The test of the small data file is included as demo data file MetaGrowth.txt. It includes 134 observations with 12 regressors. All regressors were selected as switch variable. Furthermore, the demo analysis uses standardized regressions but does not use a balanced design or specify lower and upper limits. Hence, the program was used in the slowest possible way. We used a quite old computer: a Pentium-11 computer with a low clockspeed and a 32 MB RAM, working under the operating system Windows 95. It took slightly more than 1 minute to run all 4096 model specifications. The resulting file has a size of 643 kbyte and takes a few seconds to be saved on a local hard disk. It can be loaded in about 10 seconds. On a faster computer (Pentium III, 192 MB RAM, 500 MHz), the same analysis took 19 seconds to run the regressions and 8 seconds to do the analysis.

A second test was performed with a large data file used for studying determinants of economic growth (see Florax et al., 2001 and Sala-i-Martin, 1997). This file cannot be analyzed with the demo version of the program, but the full version (see below) can handle it. The large data file is a file with 138 observations with 61 regressors. Four regressors were selected as *fixed* variable and all other variables were selected as switch variable. Furthermore, we did not use a balanced design, we used standardized regression, and we imposed an upper limit of three *switch* variables. Hence, 29260 regression equations had to be estimated. For this, we used a relatively new computer: a Pentium-III computer with a clockspeed of 500 MHz and 192 MB RAM internal memory, working under the operating system Windows 98. It took 17 minutes to run all model specifications and 75 seconds to analyze the results. The resulting file has a size of 16.932 kbyte.

### Access, modifications, and reference to MetaGrowth

We are interested to hear your opinion about the program and to receive suggestions for possible further improvement of the program. Also communication and bug reports are welcome, although we cannot guarantee to solve your problems or to fix bugs.

The downloaded program is a demo version. It contains the complete functionality of the full program, but it is not suited for dealing with large data sets (the number of regressors is restricted to 12 and the number of observations to 135). An extended version of the program is available upon request (for free) and can be obtained from Henri de Groot (E-mail: hgroot@feweb.vu.nl). Registered users will be notified when new versions of the program become available.

The program can be used free of charge. The intellectual property remains with the authors. Selling or modifying of the program is prohibited. You may use the program for your academic publications. We appreciate your referring to the program and the location where it can be found if you publish work that is based on this program.

## *Future development*

This is the first public version of MetaGrowth. Further development of the software in the future may deal with the following extensions:

- standard estimators for discrete choice models, such as logit and probit;
- weighted least squares and/or White adjusted standard errors to deal with heteroscedasticity;
- detection and exclusion possibility for outliers;
- misspecification tests, such as the White test for specification error, the Breusch-Pagan test for heteroscedasticity, and the Durbin-Watson test and Moran's $I$ for autocorrelation (temporal and spatial, respectively)
- additional output columns for the adjusted $R^2$, and various misspecification tests.

We appreciate receiving your comments on the use of the program, and suggestions for further development of the software.

# References

Belsley, D., E. Kuh and R. Welsch (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity,* Wiley, New York.

Florax, R.J.G.M., H.L.F. de Groot and R. Heijungs (2001) *Determinants of Economic Growth? A Quasi-Experimental Meta-Analytical Approach,* Dept. of Spatial Economics Working Paper, Free University, Amsterdam.

Leamer, E.E. (1983) "Let's Take the Con Out of Econometrics," *American Economic Review, 73,* pp. 31-43.

Leamer, E.E. (1985) "Sensitivity Analyses Would Help," *American Economic Review, 57,* pp. 30% 313.

Leamer, E.E. and H. Leonard (1983) "Reporting the Fragility of Regression Estimates," *Review of Economics and Statistics, 64,* pp. 306-3 17.

Levine R. and D. Renelt (1992) "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review, 82,* pp. 942-963.

McAleer, M., A. Pagan and P. Volker (1985) "What Will Take the Con Out of Econometrics?," *American Economic Review, 75,* pp. 293-307.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling (1989) *Numerical Recipes in Pascal: The Art of Scientific Computing,* Cambridge University Press, Cambridge.

Sala-i-Martin, X. (1997) "1 Just Ran Two Million Regressions," *American Economic Review, 87,* pp. 178-183; *see* also, *I Just Ran Four Million Regressions,* NBER Working Paper 6252, Cambridge, MA.