

Adjust for Windows – Version 1.1

Merz, Joachim; Stolze, Henning

Publication date:
2006

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for pulished version (APA):

Merz, J., & Stolze, H. (2006). *Adjust for Windows – Version 1.1: A Software Package to Achieve Representative Microdata by the Minimum Information Loss Principle*. (FFB-Dokumentation; No. 13). Forschungsinstitut Freie Berufe. <https://www.leuphana.de/institute/ffb/publikationen/dokumentationen.html>

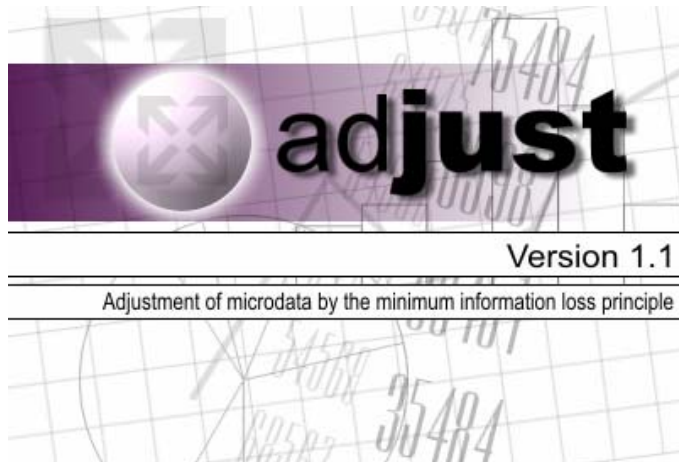
General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Adjust for Windows

Version 1.1

A Software Package to Achieve Representative
Microdata by the Minimum Information Loss
Principle

Program Manual

Joachim Merz and Henning Stolze*

* Prof. Dr. Joachim Merz, University of Lüneburg, Department of Economics and Social Sciences, Director Research Institute on Professions (Forschungsinstitut Freie Berufe, FFB), Chair ‚Statistics and Professions‘, Campus, Scharnhorststr. 1, 21335 Lüneburg, Tel.: +49 (0)4131 677 2051, Fax: +49 (0)4131 677 2059, e-mail: merz@uni-lueneburg.de, <http://ffb.uni-lueneburg.de>

ADJUST contribution: concept, algorithm and DOS-Version programming

Dipl.-Kfm. Henning Stolze, assistant at the above chair ‚Statistics and Professions‘, Research Institute on Professions (FFB), (stolze@uni-lueneburg.de; Henning.Stolze@HSI-Siteworks.de)

We thank Sascha Imme for his efforts concerning the first c++-version of ADJUST

ADJUST contribution: Windows development, design and programming

Adjust for Windows

Version 1.1

A Software Package to Achieve Representative Microdata
by the Minimum Information Loss Principle

Program Manual

Joachim Merz and Henning Stolze
FFB-Dokumentation No. 13, 2006

Summary

ADJUST FOR WINDOWS is a program package to adjust large microdata files by the information theory (entropy) based minimum information loss principle. It is useful for all that tasks, where representative results of sample data by (re)weighting are desirable. In particular, a **simultaneous and consistent adjustment** is possible with the developed nonlinear optimization approach, in which for a variety of even hierarchical ordered variables one single weighting factor per microunit is chosen to simultaneously achieve given margins after a weighted summing up.

The adjustment of microdata is an essential part of microsimulation models, too. Particularly in static microsimulation models appropriate adjustment ((re-)weighting) of a sample is the procedure to update (a starting or a resulting file) or to extrapolate the underlying sample ('static aging') into the future.

This is the program manual of the most recent development **ADJUST FOR WINDOWS**, or as a shortcut: **ADJUST/WIN**, and includes tasks, characteristics and experiences with the ADJUST/WIN package, methodological foundations and program handling, program parameters and the respective program and example files.

Running as a native 32 bit-Windows application, compared to the former versions this version benefits from a general increase in performance, unlimited microdata and restrictions (totals, margins), an easier-to-use (graphical) user interface and the ability to import microdata from various sources and database formats. **ADJUST FOR WINDOWS** is running under the following Windows versions: 95, 98, ME as well as 2000 and NT 4.0.

This documentation – besides the general handling description – includes ADJUST FOR WINDOWS updates till version 1.1.9.1.

Key words: Adjustment, (Re-)weighting, Grossing-up of samples, Survey calibration, Static Microsimulation, Entropy, Minimum information loss principle, Modified Newton-Raphson Algorithm, Windows application

Content

1	Introduction.....	5
2	The Adjustment of Microdata – Aims, Characteristics and Experiences ..	7
2.1	Aims of an Adjustment of Microdata	7
2.2	Characteristics of the Adjustment of Microdata with ADJUST	7
2.3	Practical Experiences with ADJUST to Adjust Micro (Mass)Data.....	8
2.4	Key Words.....	10
3	The Adjustment of Microdata by the Minimum Information Loss Principle - Methodological Issues	11
3.1	Sample Information, Restrictions and General Optimization Approach	11
3.2	The Adjustment of Microdata Using the Minimum Information Loss Principle ..	12
4	ADJUST FOR WINDOWS: An In-Depth User's Guide	15
4.1	Starting Adjust	15
4.2	First Look at the Program	15
4.3	Switching from evaluation to full license	16
4.4	Iteration Settings.....	16
4.5	Select the Microdata.....	18
4.6	Specify the Restrictions	19
4.7	Run the Adjustment.....	21
4.8	Runtime Error Handling	22
4.9	Saving the Resulting New Adjustment Factors.....	22
5	ADJUST Examples	24
5.1	ADJUST: Teaching Quality.....	24
5.2	ADJUST: Example of an Adjustment with Microdata of the 4th Wave of the German Socio-Economic Panel	26
5.3	To build a S-matrix manually	29
6	ADJUST: Versions.....	31
6.1	ADJUST-PC: Running on MS-DOS Or Any Compatible System.....	31
6.2	ADJUST-MICSIM: PC-Version for MICSIM	31
6.3	ADJUST-DEC: Mainframe Version for DEC	32
6.4	ADJUST-BS2000: Mainframe Version for Siemens BS2000	32
6.5	ADJUST FOR WINDOWS: Version History.....	32
7	Concluding Remarks.....	35
8	Appendix	36
8.1	ADJUST Ingredients: Sample Information, Restrictions and Iteration Settings	36
8.2	Iteration Settings.....	37
9	References	39

1 Introduction

ADJUST is a program package to adjust (grossing-up, (re-)weight) microdata (samples) by the minimum information loss (MIL) principle. It is suitable for all cases where the representativity of sample data is required by weighting or reweighting a sample, respectively a survey.

The adjustment of these microdata is an essential part of microsimulation models. Particularly in static microsimulation models appropriate adjustment ((re-) weighting) of a sample is the procedure to update (a starting or a resulting file) or to extrapolate the underlying sample ('static aging') into the future.¹

The most recent development is ADJUST FOR WINDOWS, or as a shortcut: ADJUST/WIN. Running as a native 32 bit-Windows application, compared to the former versions this version benefits from a general increase in performance, unlimited microdata and restrictions (totals, margins), an easier-to-use (graphical) user interface and the ability to import microdata from various sources and database formats. ADJUST FOR WINDOWS is running under the following Windows versions: 95, 98, ME as well as 2000 and NT 4.0.

The (former) DOS program package ADJUST, the root of the Windows version, is available in different PC- and mainframe versions: ADJUST, a stand-alone PC version is written for all IBM-compatible PCs; ADJUST-MICSIM is part of the PC microsimulation model MICSIM; ADJUST-DEC is a mainframe version for a DEC computer (1088), there is also an ORACLE version for ADJUST-DEC within the frame of the Static Sfb 3 (Sonderforschungsbereich 3 'Microanalytic Foundation of Social Policy) Microsimulation Model. ADJUST-BS2000 is a mainframe version for Siemens computer using the operating system BS2000. This version was written for requirements of the German Federal Statistical Office. The very roots of the ADJUST DOS versions are reaching back to the work of Merz within the Sonderforschungsbereich 3 'Mikroanalytische Grundlagen der Gesellschafts-politik' der Universität Frankfurt. Further developments are done at the chair of Prof. Merz at the University of Lüneburg.

The following ADJUST program manual considers the general adjustment background with focus on the new Windows version. After describing the aims, the

¹ Microsimulation in general is discussed in Merz 1991 and Orcutt, Merz und Quinke 1986; structural adjustment in static microsimulation models is the topic in Merz 1983b, 1986a,b, 1993a,b and Spahn et al. 1992.

characteristics and experiences of and with the adjustment of microdata by ADJUST the methodological issues are briefly described. Chapter 3 is an in-depth user's guide for ADJUST FOR WINDOWS. Chapter 4 is considering two examples: the first is on teaching quality with focus on an adjustment using the original survey. The second example is about data of the 4-th wave of the Socio-Economic Panel (SOEP) with focus on the sample information matrix S constructed out of the original survey by any further program package (like SPSS). Chapter 5 provides a short overview about the different ADJUST versions. The Appendix finally describes the ADJUST ingredients.

This documentation – besides the general handling description – includes ADJUST FOR WINDOWS updates till version 1.1.9.1.

2 The Adjustment of Microdata – Aims, Characteristics and Experiences

2.1 Aims of an Adjustment of Microdata

Surveys for applied socio-economic microanalytic analyses almost in general are not 'representative' for a large number of reasons like non-response, quoted data collection, errors within the field work etc. Therefore a procedure is needed for practical work to make surveys representative and to achieve representative results from the available sample.

To adjust microdata is to fit microdata to prescribed aggregate totals (restrictions, margins) by (re-)weighting. For each microunit of a microdata file (sample) a suitable weight is searched in that way that the weighted sum of all microunit characteristics will be equal to externally given aggregates (restrictions, margins or population totals).

The specific problem of a so-called simultaneous adjustment is to find for even a quite large number of a microunit's characteristics just a single weighting factor for each microunit, which after summing up fulfills all given restrictions simultaneously.

A consistent adjustment is given if for different hierarchical characteristics (e.g. households, families and persons) the desired hierarchical aggregate totals have to be achieved.

ADJUST/WIN is solving the simultaneous as well as the consistent adjustment task in an efficient and user-friendly approach for unlimited adjustment problems (except limitations due to windows memory resources).

2.2 Characteristics of the Adjustment of Microdata with ADJUST

ADJUST is solving the simultaneous and consistent adjustment problems using the minimum information loss (MIL) principle. The MIL-principle is based on a theoretical concept, the information theory. Above all, the MIL principle ensures the desired positivity condition of the weighting factors to maintain all microunits for further analyses. For the consistent solution of the adjustment problem, which simultaneously fits hierarchical microdata (e.g. households and personal information within the household and family frame), a relatively fast numerical solution with a

specific Newton-Raphson (MN) procedure by a global exponential approximation is used and developed (Merz 1994).

2.3 Practical Experiences with ADJUST to Adjust Micro (Mass)Data

With this specific MN procedure underlying ADJUST the computation efforts could be reduced by 75% for large microdata files of the Sfb 3 Microsimulation Model (with more than 60.000 households with family and personal information and with more than 250 restrictions which should be reached simultaneously). Besides the successful adjustments of large microdata files within the Sfb 3 Microsimulation Model - e.g. to analyze the German old age pension reform in the 80s - this procedure was also successfully used to adjust the Sfb 3 secondary occupation survey 1984 for the years 1990 and 2000 within Merz' analysis of the recent German tax reform 1990, investigating the impacts of the reform changes on multiple labour supply in paid and unpaid work within the formal and informal economy (Merz 1989, 1990a, 1991b). More recent applications of ADJUST at the University of Lüneburg, Research Institute on Professions (FFB, the home of ADJUST) for example involve re-weightings of the German Socio-Economic Panel with specific labour market regards of the self-employed and (liberal) professions (Freie Berufe) or adjustments of the recent FFB online survey about the start-up situation of (liberal) professions.

Besides the further own use of ADJUST for research and teaching there are now a variety of national and international ADJUST users within universities, statistical offices and the administration:

ADJUST usergroup

- Forschungsinstitut Freie Berufe (FFB), University of Lüneburg, Germany (home), Prof. Dr. Joachim Merz
- Australian Bureau of Statistics (ABS), Canberra, Australia, Steven Horn
- BIAB, Germany, Gerhard Schön
- BISS e.V., Berlin, Germany, Eberhard Pech
- Bundeskammer Wirtschaft, Österreich, Dr. Joseph Richter
- Destatis, Federal Statistical Office, Wiesbaden, Germany, Dr. Jürgen Schmidt, Dr. Erlend Holz, Dr. Manfred Ehling, Markus Zwick (e.g. adjustments of the

Time Budget Survey 1991/92 and 2001/2002, the Income and Consumption Surveys (EVS), Income Tax Surveys)

- DIW, Deutsches Institut für Wirtschaftsforschung, Dr. Stefan Bach
- ESRI, Economic and Social Research Council, UK, James Williams
- Finanzwissenschaftliches Institut, Universität Köln, Thilo Schaefer
- Fraunhofer Institut, Susanne Maidorn
- Ifo-Institut, Dr. Gebhard Flaig
- IGES GmbH, Guido Schiffhorst
- Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg, Germany, Dr. Uwe Blien
- Institut für Freie Berufe (IFB), Nürnberg, Germany, Dr. Jürgen Passenberger
- Interrogare GmbH, Gero Lipsmeier
- London School of Economics (LSE), STICERD, UK, Prof. Anthony Atkinson
- Max-Planck-Institut, Michael Bromba
- Ministry of Finance, Stockholm, Sweden, Dr. Ingmar Ericsson
- NATSEM, Microsimulation Center, University of Canberra, Australia, Prof. Dr. Ann Harding
- NIFA-Maschinenbau Panel, Sonderforschungsbereich 187, Neue Informationstechnologien und flexible Arbeitssysteme, Universität Bochum, Germany, Prof. Dr. Ulrich Widmaier
- Socio-Economic Panel, DIW Berlin, Germany, Dr. Rainer Pischner, Dr. Jürgen Schupp
- Sonderforschungsbereich 3, Mikroanalytische Grundlagen der Gesellschaftspolitik, Universitäten Frankfurt und Mannheim, Germany, Dr. Joachim Merz, Prof. Dr. H.-P. Galler
- St. Mary's University, Halifax, Canada, Prof. Andrew Harvey
- University of Köln, Prof. Dr. Schäffer, Germany
- Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim, Germany, Dr. Georg Licht

ADJUST is not only a stand-alone package, but is also a part of MICSIM, a microsimulation model for research and teaching, which is now further developed at Prof. Merz' chair at the University of Lüneburg.

2.4 Key Words

Key words: Adjustment (grossing-up, (re-)weighting) of microdata, static microsimulation, static aging, extrapolating and updating microdata, microsimulation, minimum information loss principle, modified Newton-Raphson algorithm.

3 The Adjustment of Microdata by the Minimum Information Loss Principle - Methodological Issues

A detailed discussion of the methodological foundation of the adjustment with the MIL principle and the efficient global exponential approximation is given in Merz (1983a, 1985, 1990b, 1994). Therefore, the approach is only briefly sketched below as far as it is important and helpful to run the program package ADJUST.

3.1 Sample Information, Restrictions and General Optimization Approach

Two information are necessary and have to be provided before actual running the adjustment process: first, the sample information, that is the value of all characteristics of all microunits, however and with respect to save computing time, only with those characteristics, which are essential for the actual adjustment. Within the ADJUST DOS versions this sample information matrix has to be build externally, whereas within the ADJUST Windows version, this matrix is provided by the system internally after defining the kind of restrictions (characteristics) on the logical level only.

Second, the restrictions are essential, which shall achieved as aggregate totals.

The information or sample matrix **S** encompasses the sum of all characteristics i ($i=1,...,m$) of all microunits j ($j=1,...,n$), with hierarchical microdata for a consistent adjustment. Each column (in the data file each data record) covers the information for one microunit j :

$$S_{m,n} = \left(\begin{array}{ccccc} S_{1,1} & \dots & S_{1,j} & \dots & S_{1,n} \\ \dots & & \dots & & \dots \\ S_{i,1} & \dots & S_{i,j} & \dots & S_{i,n} \\ \hline S_{i+1,1} & \dots & S_{i+1,j} & \dots & S_{i+1,n} \\ \dots & & \dots & & \dots \\ S_{k,1} & \dots & S_{k,j} & \dots & S_{k,n} \\ \hline S_{k+1,1} & \dots & S_{k+1,j} & \dots & S_{k+1,n} \\ \dots & & \dots & & \dots \\ S_{m,1} & \dots & S_{m,j} & \dots & S_{m,n} \end{array} \right) \begin{array}{l} \text{hierarchical step 1} \\ \text{(e.g. household data)} \\ \\ \text{hierarchical step 2} \\ \text{(e.g. family data)} \\ \\ \text{hierarchical step 3} \\ \text{(e.g. personal data)} \end{array}$$

In general, the problem is to find a n-vector \mathbf{p} of adjustment factors optimizing an objective function $Z(\mathbf{p}, \mathbf{q})$ - a distance function between the new adjustment factors \mathbf{p} and the available old ones \mathbf{q} - satisfying the restrictions $\mathbf{Sp} = \mathbf{r}$:

- (1) $Z(\mathbf{p}, \mathbf{q}) = \min!$
- (2) $\mathbf{S}_{(m,n)} \mathbf{p}_{(n)} = \mathbf{r}_{(m)} ..$

The objective function is minimizing the distance between the new adjustment factors \mathbf{p} that are computed and the given factors \mathbf{q} in order not to lose eventually available corrections due to non-response, sampling errors etc. within the adjustment procedure. If such corrections for a representation of the microdata were not given in advance, q_j would $1/(\text{general sampling factor})$ (or just 1) for every microunit j ($j=1, \dots, n$).

3.2 The Adjustment of Microdata Using the Minimum Information Loss Principle

The adjustment of microdata by the MIL principle is based on the information theory with the following **Information Loss**:

$$\begin{aligned} (3) \quad I(\mathbf{p}; \mathbf{q}) &= \sum_j p_j \log(1/q_j) - \sum_j p_j \log(1/p_j) \\ &= \sum_j p_j \log(p_j/q_j), \end{aligned}$$

where $\mathbf{p} = (p_1, \dots, p_n)'$,
 $\mathbf{q} = (q_1, \dots, q_n)'$ with $(p_j, q_j > 0)$,
 $\sum_j p_j = \sum_j q_j = 1, \quad (j = 1, \dots, n).$

In this procedure the information loss is that entropy based value which is given when the old adjustment factors \mathbf{q} are replaced by the new ones \mathbf{p} . Certainly there are other possible distance functions (Merz 1994). But minimizing the MIL nonlinear objective function proves to be superior to others in many dimensions; one of them is the forced positivity of the new factors to be computed.

Minimizing the information loss subject to already available adjustment factors, the nonlinear adjustment problem by the MIL principle results from:

Optimizing Problem

(4) $Z(\mathbf{p}, \mathbf{q}) = \min_{\mathbf{p}} \{ \sum_j p_j \log(p_j/q_j) \} \quad 0 < p_j, q_j < 1, \sum_j p_j = \sum_j q_j = 1,$
subject to

(5) $\mathbf{Sp} = \mathbf{r}.$

The respective Lagrangean is

(6) $L(\mathbf{p}, \boldsymbol{\lambda}) = \sum_j p_j (\log p_j - \log q_j) - \boldsymbol{\lambda}'(\mathbf{Sp} - \mathbf{r}),$
which yields a nonlinear equation system

(7) $\sum_j s_{ij} p_j \exp(\boldsymbol{\lambda}' \mathbf{s}^j - 1) = r_i \quad (i=1, \dots, m),$
that has to be solved iteratively.

The new adjustment factors with the solution $\boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ is a m -vector only, are

(8) $p_j = q_j \exp(\boldsymbol{\lambda}' \mathbf{s}^j - 1),$

where \mathbf{s}^j are the respective characteristics of the microunit j ($j=1, \dots, n$).

Restrictions as Absolute Numbers.

Usually the adjustment factors p_j and q_j are not given as relative frequencies (with $0 < p_j, q_j \leq 1$) but directly in absolute numbers with $HRF_j = p_j N$ (N is the number of all microunits of the underlying population). The absolute restrictions then are $r_i^a = r_i N$ ($i=1, \dots, m$). With given old factors

$$(9a) \quad HRFALT_j = q_j N$$

the desired new absolute adjustment factors HRF_j are

$$(9b) \quad HRF_j = HRFALT_j \exp(\lambda' s^j - 1).$$

The adjustment by the MIL principle in particular ensures the necessary positivity condition of the new adjustment factors allowing further analyses of all available microunits.

4 ADJUST FOR WINDOWS: An In-Depth User's Guide

4.1 Starting Adjust

During start-up, Adjust will check for updates, if an internet connection is available on your system. Please configure your firewall to let Adjust access the internet server of the Research Institute on Professions via the standard http-port (8080). If updates are available, a message window will appear offering a link to download an updated setup-file.

4.2 First Look at the Program

Run ADJUST FOR WINDOWS by double-clicking the appropriate program file or locate the corresponding entry in the Windows program menu (click the "Start"-Button and select "Programs").

After ADJUST has been loaded, you will face a screen similar to the one below:



Although the steps necessary to run an adjustment are similar to those of former versions, the way to do things is completely different. The text mode has been replaced with a standardized Windows look-and-feel interface.

To trigger the key functions of ADJUST FOR WINDOWS, you have to click the appropriate button.

“About”, “Help” and “Exit” should be self-explanatory.

The button “Iteration settings” displays a screen where you can set the parameters you would have defined – if you would not take the default values - in a file called ADJUST.INF in earlier versions (cf. chapter 3.2).

Clicking the button “Microdata” presents you with a Windows dialog where you have to select the data source containing the microdata (cf. chapter 3.3). Note: To lower the possibility of usage errors, the following options become available only when they are needed. This step-by-step approach will ensure that no step necessary in order to properly run an adjustment process will be (or can be) skipped.

After selecting a data source containing the microdata, you will have to select a source containing the restrictions. Click the button “Restrictions” (cf. chapter 3.4) and proceed to the next step.

After having selected both the microdata and the restrictions, you are ready to run the adjustment process. Click “adjust!” and observe the progress (cf. chapter 3.5).

After the adjustment is finished and you return to the main screen, you can see statistics and save the results that were calculated during the adjustment process.

Click “Save results” and select where you want to save your results (cf. chapter 3.6).

After this brief overview we discuss the single steps in more detail.

4.3 Switching from evaluation to full license

If you have obtained a full license of Adjust, you can enter the licensing menu by pressing the ‘L’-key on your keyboard. A small window with a text field will appear where you can enter your license code to unlock and to get the full feature set of Adjust.

4.4 Iteration Settings

This dialog allows you to set the different options that influence the adjustment process. The default values are the same as those used in the (now obsolete) file ADJUST.INF (see the example in chapter 5 and the Appendix). In most cases, the default values will work fine with any adjustment. As the computing power of

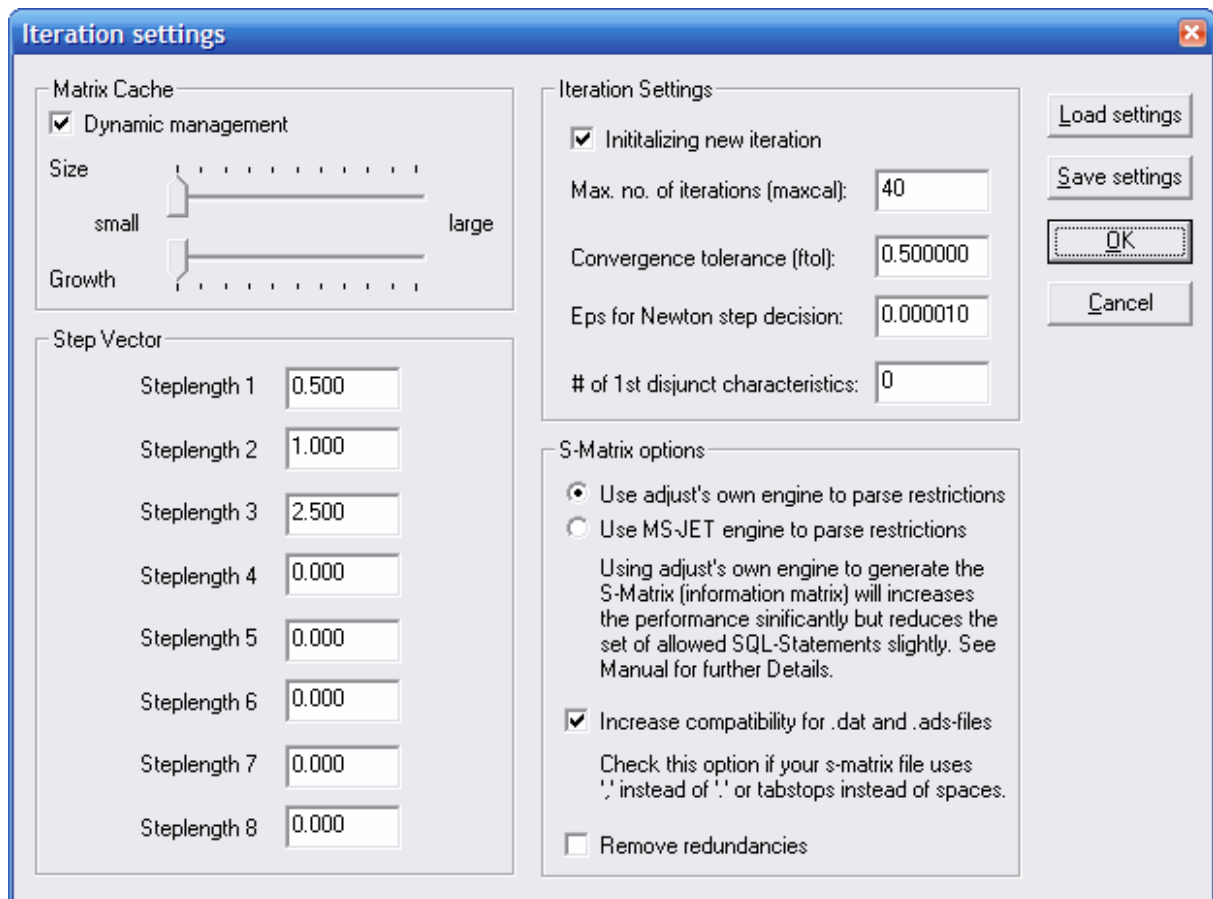
standard PCs significantly increased during the last years, fine tuning for performance improvements won't be necessary.

However, in Version 1.1 there are a few new features. One is to burst the performance when generating a new S-Matrix from an Access-database. Using the standard SQL-Parser, which is included in the Microsoft Jet engine used by Adjust, is sometimes slow. This depends heavily on your system configuration and is beyond our influence. Especially when working with many cases or restrictions in your database, this process may take from some minutes up to several hours. Instead, you can use Adjust's own engine to parse your restrictions to improve the performance significantly. However, the SQL command set is limited to standard comparison operations (<, <=, =, >=, >, != [not equal]) and the logical operators AND [logical and] and OR [logical or] as well as an unlimited number of parentheses.

Another new feature is the option to increase the compatibility of S-Matrix files. When enabled, you may use tab stops instead of spaced to separate your variables or commas as decimal delimiter. Additionally you may choose to remove redundancies (restrictions with no corresponding cases). Processing this option at the beginning of the adjustment process may take some time depending on the size of your dataset and the configuration of your computer. It is recommended to avoid such restrictions in advance.

Besides setting the preferences for the adjustment process, you can also load existing settings or save them for later use. Just click the appropriate button and enter a new filename or select an existing file from the file select dialog.

Note: The matrix cache is currently optimized for maximum performance. Fine-tuning the settings is currently disabled because of already efficient runtime performance. Therefore any changes you make to the matrix cache settings are not reflected in the program. Future versions of ADJUST FOR WINDOWS may include support for fine-tuning.



4.5 Select the Microdata

ADJUST FOR WINDOWS supports more than just ASCII-files as data sources. The current version allows you to use basically three different file formats as data sources:

The old file format required by former versions of ADJUST. (Select DOS-Adjust data files “.dat”)

An improved version of format 1 (Adjust matrix files “.ads”). It is now possible to import data in a less strict way. Instead of exactly formatting the data with spaces (as required by format 1), you can now use output generated by almost every database application. No need to manually fill up gaps with additional spaces. (Note: ADJUST FOR WINDOWS natively generates microdata files in this format.)

Use a Microsoft Access database as a data source (“.mdb”): ADJUST FOR WINDOWS supports standardized database interfaces like ODBC and DAO. Future versions may add support for many more database application specific formats.

According to the format of the input file you selected, the way of specifying the restrictions will be different (cf. chapter 3.4).

4.6 Specify the Restrictions

As mentioned before, ADJUST FOR WINDOWS supports various file formats as the data source for the microdata. If you have the microdata in the old “.dat” or the new “.ads” format, you just need to give the name of the file containing the restrictions. The file can be either in the old format used by former versions of ADJUST (“.dat”) or in the new “.adr” format generated by a previous run of ADJUST FOR WINDOWS (see below).

If the data is supplied via a Microsoft Access database file, you will be presented with the interface below when clicking the button “Restrictions”. This is a powerful editor for working with sets of conditions. You can load and save a set of *logical* restrictions (e.g. Einkommen < 100000; stored as a “.adr” file), and you can edit existing sets with ease.

Create new sample matrix

General Settings

Identification field: ID

Old adjustment factors: HRF

Save adjust-type matrix and restriction files later for further use: C:\Adjust

OK, Cancel, Load, Save

Restrictions

Alter <21 -> 1113

Delete

Available variables:

- Alter
- Besuchshäufigkeit
- Geschlecht
- HRF
- ID
- Stoffvermittlung-Plenar
- Stoffvermittlung-Tutoriel
- Stoffvermittlung-Vorlesung
- Studiengang

Use standard SQL-Statements like [variable] <>= [value] ... AND ... OR ... to define this restriction:

Alter <21

Number of cases to be achieved: 1113

Update, Add, Add & Clear, Clear

SQL-Help

HSI LRE Interface Rev. 3

Display microdata

The general settings

You have to specify both the field containing the ID and the field which holds the old adjustment factors. The drop-down lists contain all fields within the table you selected after opening the database file in step 1 (cf. chapter 3.3). The field below allows you to choose the filename and the location of the resulting “.ads” and “.adr” files. In the

above example, the microdata will be written to the file "J:\Adjust1.ads", and the restrictions resulting from your input in the Logical Restriction Editor will be written to the file "J:\Adjust1.adr". This allows you to distribute your data in a more easily readable form, and enables you to later re-run exactly the same adjustment (maybe with different iteration settings) by simply specifying the ".ads" and ".adr" file.

Working with sets of restrictions

The drop-down list contains the restrictions you already defined. When you select an entry from the list, the fields below fill with the existing values. You can now edit this existing entry or add a new entry based on the existing values. If you want to remove the selected list entry from the set of restrictions, just click the button "Delete".

Editing restriction sets is quite easy and straightforward. In the above example, 234 cases should have an income ("Einkommen") less than 100,000 (0.1 Million in the currency of the selected data source). The purpose of the four "edit buttons" is described below.

Update: This updates the selected list entry with any changes you made. If, for instance, you want to change the number of cases that are to be achieved from 234 to 230, you would just type in 230 in the appropriate field (there are just two fields in which you can type directly) and click the button "Update". The change is reflected instantly in the list.

Add: If you click the button "Add", a new entry is added to the list. So clicking "Add" in the above example will create a double entry.

Add & Clear: This button does basically the same as the button "Add". The difference is that the two fields where you enter the restrictions-specific information are emptied. This allows faster creation of a new set of restrictions because you avoid manually deleting the contents of the text field.

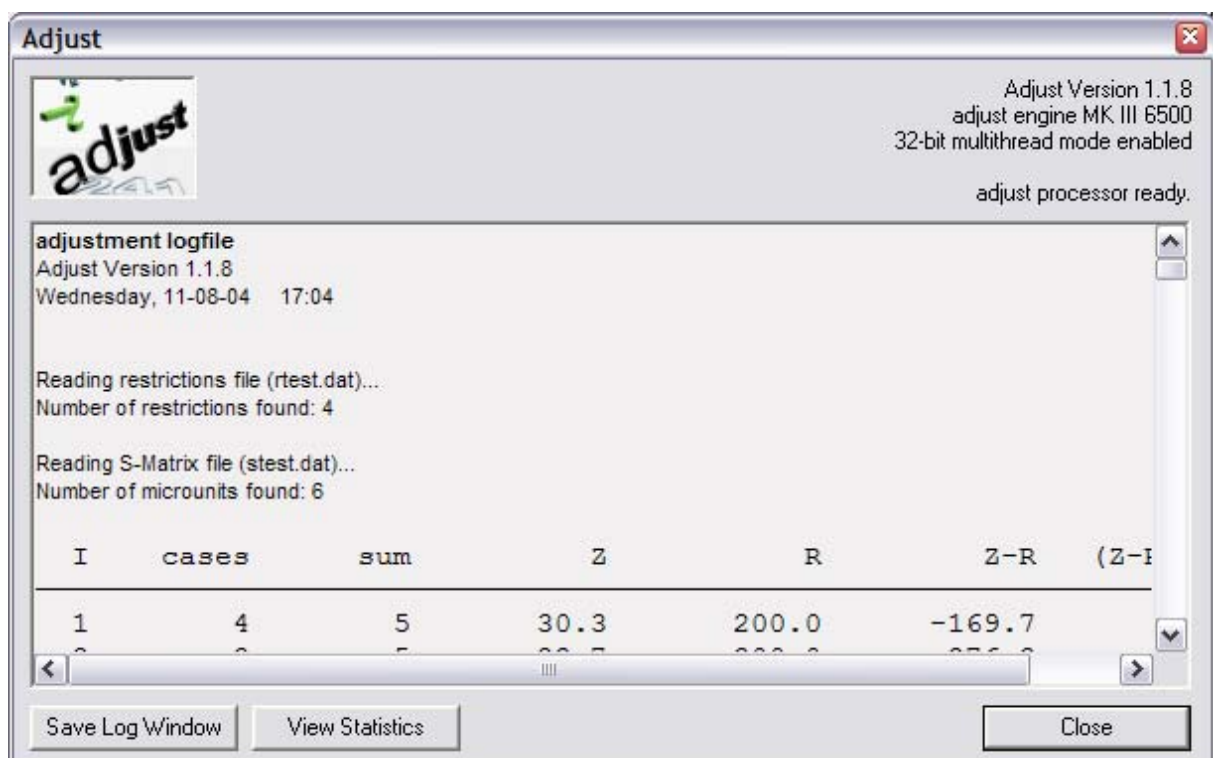
Clear: Clicking on this button simply empties the two editable fields.

The list “**Available variables**” holds all field names in the selected database table. It is not only a reference for you to know which variable names you can use, it is also a possibility to save time by double-clicking an entry in this list: For instance, double-clicking the entry “Einkommen” in the above example will add this variable name to the field containing the definition of the restrictions (10 times double-clicking a list entry is usually faster than 10 times typing a 9-letter word).

The button “Display microdata” does exactly what its name implies. A new window will be opened showing you the contents of the selected database. So it is basically a database viewer.

4.7 Run the Adjustment

After the necessary pre-adjustment steps are done, you can click the button “adjust”. A dialog similar to the one below will open and will inform you about the proceeding of the adjustment process.



The whole process is often complete in just a few seconds. Most of this dialog contains the logfile information. All steps and all intermediate results are written and explained in this logfile. Since it is usually a good idea to keep the logfile information for future reference, you can click the appropriate button and save the information to a standard text file editable with every text editor.

Clicking the button “Close” will close the log window and return you to the main screen where you can save the results that were calculated during the adjustment process.

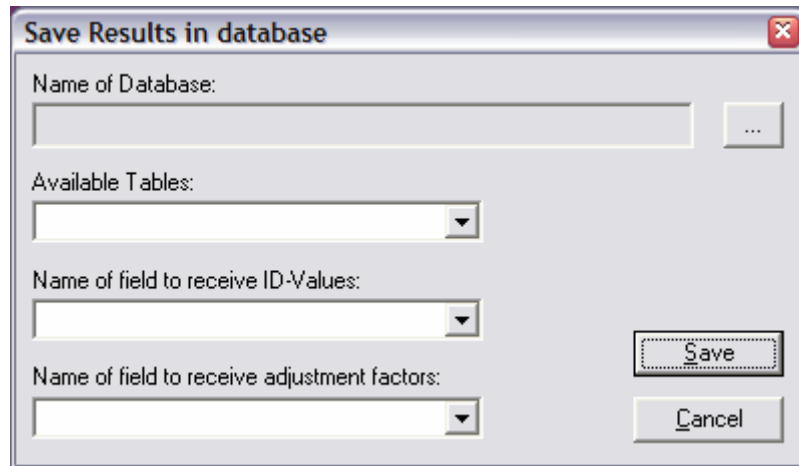
To get some more detailed information about the adjustment process, the results and their quality, you can click the button “View statistics”. You can then switch between a graphical comparison between the old and the new factors, and a textual analysis of the results with various statistically relevant values (these statistics can be saved by clicking on the button labeled “Save...”).

4.8 Runtime Error Handling

The underlying optimization algorithm is very robust and has proved for reaching final convergence based on theoretical reasons and based on numerous applications. If there is any runtime error, then the reason *always* is the impossibility to invert a needed matrix, i.e. substantially there is at least one linear dependent restriction. For instance, if the restrictions – and therefore one row of the sample matrix **S** – includes the number of men, women and the total population (possibly indirectly defined by other restrictions), then one of this information is obsolete. Without loss of accuracy skip one of these characteristics and the algorithm will converge (we bet you for one box of fine sparkling wine!).

4.9 Saving the Resulting New Adjustment Factors

After a successful run of the adjustment process, you may want to save the results, the new adjustment factors. When you click the appropriate button, you will be asked where you want to save your results, resp. in which format. You can choose between a text file (in this case all you have to give is a filename) and a database. If you choose the latter, you need to give some information about how to store the results in a database (see the example below).

A screenshot of a Windows-style dialog box titled "Save Results in database". The dialog box has a standard title bar with a close button (X) in the top right corner. Inside the dialog, there are four input fields, each with a label and a dropdown arrow on the right side. The first field is labeled "Name of Database:" and has a small "..." button to its right. The second field is labeled "Available Tables:". The third field is labeled "Name of field to receive ID-Values:". The fourth field is labeled "Name of field to receive adjustment factors:". To the right of these fields are two buttons: "Save" and "Cancel".

First of all you need to give the name of the database file. If the file does not exist, it will be created.

Then you need to specify a table in which to save the results. You can either add new fields to an existing table or you can add a new table to the database by selecting the corresponding entry as shown in the above example.

Finally, you have to choose the names of the fields in which the resulting new adjustment factors **p** and the corresponding ID-Values are to be saved. Although it is possible to add the results to an existing field, it is recommended that you create new fields or a new table for your results. Separating them clearly from other data might avoid confusion. Click "Save" and the database will be created / updated.

5 ADJUST Examples

Let us consider two examples: the first is on teaching quality with focus on an adjustment directly using an original survey. The second example is about data of the 4-th wave of the Socio-Economic Panel (SOEP) with focus on a sample information matrix **S** constructed out of the original survey by any further program package. You will find both data sets on the ADJUST CD.

5.1 ADJUST: Teaching Quality

The following teaching quality example will demonstrate the powerful features of ADJUST FOR WINDOWS's all-new database engine. It will allow the user to access a database containing the original record sets of the survey to be adjusted directly (the **S** information matrix is built-in and handled internally and automatically) and to design the restriction on the logical level only.

In this example, 20 students at the University of Lüneburg were asked about the quality of lectures and tutorials. Additionally they were asked to give some sociodemographic information (age, gender, study subject etc.), which are available for a demographic adjustment. The necessary aggregates were obtained from the University enrollment office and can be considered as reliable 'restrictions'.

The following Table 1 is an extract of all survey information showing here only some demographic data; Table 2 thereafter displays the aggregate data (restrictions, margins, totals) provided by the University enrollment office to be achieved.

Table 1: Survey data

ID	old	weighting	gender	subject	age	...
1	100		true	0	21	
2	100		true	1	25	
3	100		true	2	20	
4	100		true	0	23	
5	100		true	3	29	
6	100		false	2	24	
7	100		false	0	19	
8	100		true	2	22	
9	100		true	2	23	
10	100		false	1	32	
11	100		true	0	26	
12	100		false	1	20	
13	100		false	0	19	
14	100		false	0	21	
15	100		false	2	23	
16	100		false	3	24	
17	100		false	0	26	
18	100		false	0	20	
19	100		true	2	23	
20	100		false	0	24	

Table 2: Restrictions: logical condition and absolute number of students

Logical Condition	Aggregate number
age < 21	1113
age > 20 AND age < 25	2187
gender = TRUE	2314
subject = 0 AND gender =	837
subject = 0 AND gender =	674
subject = 1 AND gender =	344
subject = 1 AND gender =	262
subject = 2 AND gender =	441
subject = 2 AND gender =	1131
subject = 3 AND gender =	247

After having collected the questionnaire information as a Microsoft Access Database File, the preparations for the adjustment process are very simple and designed straight-forward (Note, if other data formats are required (like from SPSS etc.) and might not be readable by ADJUST FOR WINDOWS just import these data into a MS Access Database File). First of all, the database file has to be opened and the table containing the data has to be specified. After that, clicking on the ‚Restrictions‘-button will cause the program to open the Logical Restriction Editor. Within this tool, you can define the conditions of the restrictions using simple SQL-Statements like ‚gender = true‘ or ‚age < 24 AND gender = true‘ and specify the aggregate numbers, e.g. for

,gender = true' 2314 is the total number of female students enrolled at the University of Lüneburg in 1999.

On the user's wish, the internally calculated **S**-matrix and the restrictions-file can be stored on the hard disk for further use.

After the adjustment process has been initialized by clicking the ,adjust'-button in the main dialog window, the new weighting factors can be saved directly to the database file or as separate MS Access- or textfile.

As you can easily see, it is not necessary to prepare your data with any other programs. Just let ADJUST FOR WINDOWS access your database and specify your conditions and aggregate numbers obtained from a reliable source. The adjustment of microdata has never been that easy and efficient before.

5.2 ADJUST: Example of an Adjustment with Microdata of the 4th Wave of the German Socio-Economic Panel

The following example is based on a subfile with 760 household records from the 4th wave of the German Socio Economic Panel. The structure of the incorporated 8 variables is given in Table 3.

Table 3: Structure of the Example-Subfile of the 4th Wave of the German Socio-Economic Panel

HHNRAKT/*	Household (HH)-No.	*/ NOT NULL	NUMBER (5)
GES /*	Sex of the HH-Head (H)	*/	NUMBER (1)
GEB /*	Year of birth of the HH-(H).	*/	NUMBER (3)
FAM /*	Family status of the HH-(H).	*/	NUMBER (1)
HEK4 /*	Net income of the HH	*/	NUMBER (5)
BG4 /*	Gross income of the HH	*/	NUMBER (5)
AZP /*	No. of persons in the HH	*/	NUMBER (2)
HRF /*	old adjustment factor	*/	NUMBER (9,2)

Table 4 now defines the substantial or logical restrictions and their aggregate numbers (restrictions) to be achieved by the MIL-procedure.

According to Table 4 the example sample information matrix **S** includes for each microunit j $m=10$ restriction characteristics (and additionally the id: HHNRAKT and the old factor HRF) (stored in *smicsim0.dat*, respectively built internally within the Windows version). The respective aggregate numbers **r** are stored in the restriction file *demo2.dat*.

It has to be noted, that e.g. by this sample the last restriction characteristic is sparsely occupied (there is only one household with this characteristic in *smicsim0.dat*). Even for such an extreme 'biased' data basis ADJUST is able to find a final solution.

Finally, the iteration settings for this example might be the following, which are summarized in the file *adjust.inf*. Within the Windows version these iteration settings are the default values provided by ADJUST/WIN.

Table 4: Given Aggregates for the Example: Restrictions of the 4th Wave of the German Socio-Economic Panel

Substantial Restrictions	Restrictions (Number of Households)
married male household head, born between 1946 and 1956 fam=1 and ges=1 and geb >946 and geb<=956 ¹⁾	3.000.000
married male household head, born between 1936 and 1946 fam=1 and ges =1 and geb>=936 and geb<=946	3.500.000
married household head, household size>4 fam=1 and azp>4	1.300.000
married household head, household size>5 fam=1 and azp>5	250.000
married household head, gross income of household>6000 fam=1 and bg4>6000	3.800.000
unmarried, male household head, gross income of household>6000 fam!=1 and ges=1 and bg4>6000	300.000
unmarried, female household head, gross income of household>6000 fam!=1 and ges=2 and bg4>6000	300.000
married female household head fam=1 and ges=2	400.000
married household head, household size>2 fam=1 and azp>2	1.600.000
single and household size>4 fam=2 and azp>4	10.000

1) respective SQL-command

adjust.inf:

demo2.dat	NAMRES
10	NNP
smicsim0.dat	NAMFIL
760	NMU
0	ISTART
40	MAXCAL
1	IPRINT
0.5	FTOL
0.00001	EPSHL

1	N1
T	LHRF
newhrf.dat	NAMHRF
3	NNL
0.5	STEP(1)
1.0	STEP(2)
2.5	STEP(NNL)

To summarize: in this example the $m=10$ restrictions are given in *demo2.dat*, the information matrix **S** with $n=760$ microunits (households) - out of a subfile of the German Socio Economic Panel - is given in *smicsim0.dat*. Within the adjustment process there are three alternative steplenghts for each iteration, where Step (NNL)=2.5 is the maximum allowed value of the global exponential approximation with iteration dependent steplenghts. The new adjustment factors are automatically written in a new file called *newhrf.dat* where the first column encompasses a key (here household number HHNRAKT) for a further merge with the data of the above subsample or an entire sample.

5.3 To build a S-matrix manually

The S-Matrix file contains the transposed S-matrix as well as two additional information for each micro unit: An identifier and the old weight. So each line of the .ADS-file represents one case: the first column contains the identifier, the second the old weight and the following the logical information characterising the corresponding micro unit according to the aggregate data. The values are separated by spaces or semicolons; the end of a micro unit is terminated by a linefeed.

(ID) (old weight) (information on restriction 1) (information on restriction 2) ...

This logical information can be easily generated with common statistic programs (e.g. in SPSS with the "recode to new variable"-command).

Additionally you will need an .ADR-file containing the aggregate values for the restrictions to be achieved. These values are listed in one column in a text file, separated by linefeeds:

(Aggregate sum of restriction 1)

(Aggregate sum of restriction 2)

...

Example: Person 1 is 30 years, person 2 is 40 years of age. After the recalibration, the sample should represent 250 persons under the age of 35 and 150 persons older than this. The old weights were all 200. The needed S-matrix looks like the following:

1	200	1	0
---	-----	---	---

2	200	0	1
---	-----	---	---

With a restriction file of the following content:

250

150

6 ADJUST: Versions

The program package ADJUST under DOS is available in different PC- and mainframe versions. The calling sequence and working method of the different versions are in principal identical. ADJUST FOR WINDOWS, however, has a fundamentally different design and programming.

Based on the MIL-algorithm used in the former ADJUST-PC Fortran written version, this new version is enhanced, redesigned and completely rewritten in C++. Running under any Win32 platform, ADJUST FOR WINDOWS is the most recent and most advanced development in the series of ADJUST versions. ADJUST FOR WINDOWS is running under all actual Windows versions: 95, 98, ME, 2000, NT 4.0. It has virtually no restrictions according to the number of aggregates to be achieved and the underlying size of the sample.

6.1 ADJUST-PC: Running on MS-DOS Or Any Compatible System

ADJUST-PC - the stand-alone DOS PC version - was written for all IBM-compatible PCs. The respective FORTRAN 77 compiler is a Microsoft compiler.

Hardware requirements:

- 640 KB,
- 286 and above versions,
- if possible mathematical coprocessor.

6.2 ADJUST-MICSIM: PC-Version for MICSIM

ADJUST-MICSIM is in particular a part of the PC microsimulation model MICSIM (see Merz and Buxmann 1990, Merz 1993). The working method corresponds to ADJUST-PC. The construction of the information matrix **S** and the restrictions **r**, in particular, are user-friendly and interactively supported. The construction has to be done only on a substantial level giving just the definitions like in Table 4. MICSIM then computes the respective information matrix **S** and provides this information in the corresponding ASCII file for efficient computation with regard to the iterative process.

6.3 ADJUST-DEC: Mainframe Version for DEC

ADJUST-DEC is the mainframe version for a DEC computer (1088) with an appropriate FORTRAN 77 compiler.

Within the framework of the Static Sfb 3 Microsimulation Model (Merz 1993) there are two subversions: the first subversion uses the relational database system 1022, the other subversion is based on the relational database system ORACLE, which is handling the I/O via respective SQL-interfaces.

6.4 ADJUST-BS2000: Mainframe Version for Siemens BS2000

ADJUST-BS2000 is a mainframe version, which is particularly developed for the requirements of the German Federal Statistical Office in Wiesbaden. Among others ADJUST-BS2000 was used for the adjustment of micro data from the nation-wide Time Budget Survey.

6.5 ADJUST FOR WINDOWS: Version History

This is an overview of the improvements in Adjust's sub-versions so far. If there is internet access during the start-up of Adjust, a check for updates will be performed and you'll be given the opportunity to download the newest version, if there are updates available.

New in version 1.1.9.1:

- Increased compatibility with SPSS when writing adjustment results to a textfile. New factors are now written in an enhanced text format. You can choose between points or commas as decimal separator.

New in version 1.1.9.0:

- Added support for floating-point-variables in the S-matrix. This may slow down the adjustment process a bit, depending on your system. For special binary or integer builds optimized for very large data files, please contact ffb@uni-lueneburg.de

New in version 1.1.8.9:

- Added support for MS Access database files up to Access 2003. The Setup Program will install and register the dao360.dll (version 3.6) if not already installed on the system.

New in version 1.1.8.7:

- Fixed bug accessing the registry for the new license model. Administration rights are required for the first start after the setup and for upgrading to the full version, but no longer for normal operation.
- Added new error handling (error #17). If the distance function now reaches infinity due to linear dependencies in the data the optimization will stop and an error message will appear.

New in version 1.1.8.5:

- Automatic update function added.
- Increased compatibility mode for s-matrix-files. A checkbox for this option is available in the preferences dialog.
- Option to remove redundancies (restrictions with no corresponding cases) from the memory. This option is helpful when dealing with large databases. You can enable this feature in the preferences dialog.

New in version 1.1.8.0:

- Improved performance with large data files.
- Removed restrictions concerning disjunct characteristics.
- Added a new license-system. You are now able to switch from the evaluation copy to the full license by pressing the 'L'-Key in the main Windows and entering your personal license code.
- Adjust is using a new Installer. DAO 3.5 runtime libraries are now installed automatically during the installation process.
- Fixed some minor bugs with MS Access-files.

New in version 1.1.5.0:

- Added native SQL parsing to improve performance when using large data files. Native mode is standard now. To switch of this option and to use the complete sql-language command set see the preferences dialog.
- S-Matrix files can now store floating-point values.
- Fixed minor bugs.

7 Concluding Remarks

Please, let us know if there are any difficulties or possible improvements in running ADJUST. We do very appreciate your comments and certainly do want to help.

Enjoy your work!

Address:

Prof. Dr. Joachim Merz, University of Lüneburg, Department of Economics, Behavioural and Law Sciences, Research Institute on Professions (FFB), Campus Scharnhorststr. 1, 5/32, 21335 Lüneburg, Germany.

Tel.: +49 (0) 4131 677 2051; Fax: +49 (0) 4131 677 2059

e-mail: ffb@uni-luenebrug.de

internet: <http://ffb.uni-lueneburg.de>

8 Appendix

8.1 ADJUST Ingredients: Sample Information, Restrictions and Iteration Settings

Sample Information and Restrictions (Aggregate Data)

As described in the first chapter there are **two data sources** which are essential: the sample information matrix **S** and the restrictions **r**.

In the ADJUST DOS versions **S** has to be constructed in advance. The restrictions can be set in the beginning of the actual ADJUST run. In the ADJUST FOR WINDOWS version **S** is constructed internally. If **S** is already given, ADJUST FOR WINDOWS is using this information, too; thus it is compatible to the former DOS-Versions. The user only has to define the restrictions on a logical level and to enter the absolute restriction numbers (**r**). Additionally, the user can store the actual **S** and **r** matrices for further runs.

S-Matrix:

The **S**-matrix provides the sample information with respect to the single restriction types. The rectangular source consists of n records. For each microunit j ($j = 1, \dots, n$) a record incorporates the microunit's identification number, the old adjustment factor and all further m restriction characteristics (in so far, the entire file is the transposed **S**) as:

IDREC, HRFALT, (S(I,J), $I = 1, \dots, M$)

where

IDREC = microunit j identification (e.g. household number) (INTEGER)

HRFALT = old, given adjustment factor of microunit j (INTEGER)

$S(I,J), I = 1, \dots, M$ = column vector s^j of all m adjustment characteristics (REAL)

Restrictions:

The second data source contains the m restrictions $r_i (i=1, \dots, m)$ in one single record.

8.2 Iteration Settings

Within the ADJUST FOR WINDOQWS version, iteration settings are directly available as default values, which might be changed ad libidum.

The ADJUST DOS versions expect to have input/output-information first. This option generally allows to run the package from the screen or via a prepared BATCH file. The output will be given accordingly, either on the screen or to a file.

If the information is expected via the screen, the following interactions with a respective input information are necessary:

ISTART	(=0 initializing, =1 current iteration): Normally ISTART is = 0, if former iteration results are used.
MICSIM	(ADJUST via inf-file =1, interactive =0): If the following information is given in an information file called ADJUST.INF than insert 1.
NAMRES	(filename of restrictions (r)):
NNP	(number of restrictions):
NAMFIL	(file name of sample matrix (S)):
NMU	(number of microunits):
MAXCAL	(max no. of iterations each run):
IPRINT	(print each IPRINT iteration):
FTOL	(convergence tolerance):
EPSHL	(eps for Newton step decision): gives the tolerance of different variable steplength in comparison to the Newton steplength (=1): \leq EPSHL, than the Newton steplength is used.
N1	(no. of first disjunct characteristics):
LHRF	(logical switch for writing HRF-FILE): Is LHRF=T than
NAMHRF	(name of new HRF-file):
NNL	(number of steplengths): Here alternative steplengths can be entered for the iterations in addition to the automatic iteration-dependent-global-exponential approximation of an optimal steplength. One of these steplengths should be one to compute also the normal Newton-Raphson procedure. Attention: the last value of the steplength is the maximum value for the global

exponential approximation. For the following iterations, ADJUST uses the steplength which provides greatest success.

STEP(...) (vector of steplengths)

NOTE: STEP(NNL) = EPSTEP (=max value of the exponential approximation)

9 References

- Merz, J. (1983): Die konsistente Hochrechnung von Mikrodaten nach dem Prinzip des minimalen Informationsverlustes, Allgemeines Statistisches Archiv, 67, 342-66.
- Merz, J. (1985): Ein modifiziertes Newton-Verfahren zur Lösung des Hochrechnungsproblems nach dem Prinzip des minimalen Informationsverlustes, Computing, 35, 51-61.
- Merz, J. (1986): Structural Adjustment in Static and Dynamic Microsimulation Models, Orcutt, Merz und Quinke (1986), 423-46.
- Merz, J. (1986): Das statische Sfb 3-Mikrosimulationsmodell - Konzeption und Realisierung mit einem relationalen Datenbanksystem, Angewandte Informatik, 5, 205-12.
- Merz, J. (1991): Microsimulation - A Survey of Principles, Developments and Applications, International Journal of Forecasting, 7, 77-104.
- Merz, J. (1993): Das statische Sfb 3 - Mikrosimulationsmodell - Mainframe- und PC-Version. Hauser, R., Ott, N. and G. Wagner (Ed.): Fortschritte in der Wirtschafts- und Sozialberichterstattung - Methodische Beiträge, Weinheim.
- Merz, J. (1993): Microsimulation as an Instrument to Evaluate Economic and Social Programmes, FFB-Discussion Paper No. 5, Department of Economics and Social Science, University of Lüneburg, Lüneburg.
- Merz, J. (1994): The Adjustment of Microdata Using the Minimum Information Loss Principle, FFB-Discussion Paper No. 8, Department of Economics and Social Science, University of Lüneburg, Lüneburg.
- Merz, J. and P. Buxmann (1990): MICSIM - A PC-Microsimulationmodel for Research and Teaching, realized with C and the relational Databanksystem ORACLE, Sfb 3 -Working Paper No. 316, Frankfurt/Mannheim.
- Orcutt, G.H, Merz, J. and H. Quinke (Ed.) (1986): Microanalytic Simulation Models to Support Social and Financial Policy, North-Holland, Amsterdam.
- Spahn, P.B., Galler, H.P., Kaiser, H., Kassella, T. and J. Merz (1992): Mikrosimulation in der Steuerpolitik, Physica-Verlag, Berlin.