SFB 649 Discussion Paper 2007-061

Embedding R in the Mediawiki

Sigbert Klinke*
Olga Zlatkin-Troitschanskaia**

* Humboldt-Universität zu Berlin, Germany
** Johannes-Gutenberg-Universität Mainz, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Embedding R in the Mediawiki

Sigbert Klinke[1], Olga Zlatkin-Troitschanskaia[2]

[1]Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät, Institut für Statistik und Ökonometrie, Spandauer Strasse 1, 10178 Berlin, Germany, sigbert@wiwi.hu-berlin.de
[2]Johannes-Gutenberg-Universität Mainz, Fachbereich Rechts- und Wirtschaftswissenschaften, Lehrstuhl für Wirtschaftspädagogik, Jakob-Welder-Weg 9, 55099 Mainz, Germany

Key words: R, wiki, Mediawiki
JEL Classifications: C63

1 Introduction

Teaching statistics to students in our area of economics and educational science often brings about the problem that students have either forgotten their statistical knowledge, or have taken different classes than the ones we offer in basic statistics. We therefore need some kind of statistical dictionary where we, as teachers, can refer to a common base and where students can look up specific terms. The Wikipedia - a general online encyclopaedia - compelled us to use a wiki for our dictionary. While the Wikipedia contains a large number of statistical terms, these are often too long and detailed to be visual displayed in lectures very well and some more specific terms are not included.

Similar projects to ours are already available in German, namely:

1. *ILMES - Internet lexicon of methods for empirical social science* by Ludwig-Mayerhofer (2004) and
2. *Statistics A-Z* of the german Federal Statistical office.

[0]This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

These are both based on pure HTML coding and can only be modified by their owners.

Once we had decided to use a wiki as our platform, we needed to find a way of somehow integrating statistical graphics and tables. This led us to the idea of embedding R into our wiki and thereby creating another way of using R as defined by Leisch (2007).

Below we will:

- Describe which wiki we chose.
- Show some simple examples of what can be done with the extension.
- Briefly describe some wikis which use the extension.
- Provide a forecast of what the next steps might be.

2 Which wiki is the best?

Every suitable wiki needs to fulfill some basic requirements, these are:

1. The integration of graphics
2. The integration of complex tables
3. The integration of formulas, preferably written in LaTeX
4. Support for Unicode and multiple languages including right-left languages for later projects
5. Export to PDF or XML to enable printouts.
6. The integration of graphics and tables generated from statistical software (in this case R).

The first five conditions considerably limit the choice of the wiki (see Table 1). None of the wikis naturally R and they all require an extension to be written.

We therefore decided to use Mediawiki, because:

Wiki	Math formula	Multiple lang. and Unicode support	Complex tables	PDF and XML export
DokuWiki	plugin	yes	no	none
JSPWiki	plugin	no	yes	PDF
Mediawiki	yes	yes	yes	yes
MoinMoin	plugin	yes	yes	yes
MoniWiki	plugin	no	yes	XML
Oddmuse	plugin	yes	yes	none
PhpWiki	plugin	no	yes	PDF
PmWiki	plugin	no	yes	yes
TiddlyWiki	plugin	yes	yes	none
TikiWiki	plugin	no	yes	PDF
TWiki	plugin	no	yes	yes

Table 1: Overview about some selected wikis and their properties from `www.wikimatrix.org`. Note that the quality of the mathematical formulas or the PDF export might be sub-standard!

- The Wikipedia is based on Mediawiki; this should therefore guarantee future development.
- It allows us to transfer high quality student work to the Wikipedia or to the Wikibooks.
- It has all the desirable features.

The DokuWiki, the wiki behind the R Wiki, could have been chosen, but it still lacks some features. From an administrative point of view, DokuWiki's approach of storing the wiki data in one file is very appealing; in contrast, Mediawiki's approach of using a MySQL databases and the file system itself creates severe problems when moving a wiki to another location. Moreover, the R Wiki developers plan to integrate the R code by forwarding and executing it on a separate machine; this seems much too complex for our purposes.

Our main aim is to produce online documents with embedded tables and graphics; like Sweave (Leisch 2002) for LaTeX. Documents where the user

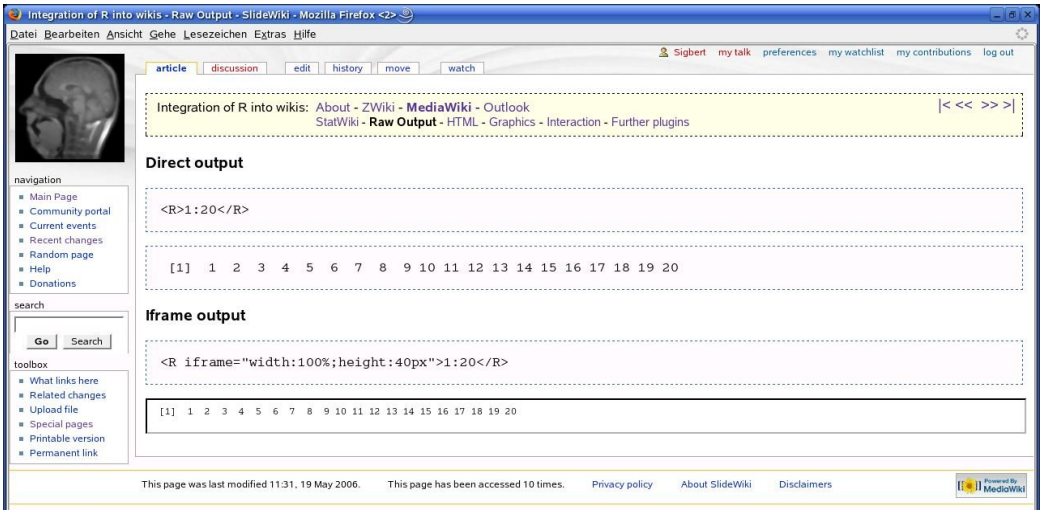

Figure 1: Raw output of the R program `1:20` in the WebPresentation wiki.

can also interactively program with R are beyond our scope, because only a few of our students use R.

In order to generate the desired graphics and tables from R we wrote a PHP extension for Mediawiki, which adds several tags and attributes to embed R programs in the wiki pages. The R programs are an integral part of the wiki page and are as accessible as the wiki page itself. Experience has taught us to usually just limit access to the wiki pages to registered users only in order to avoid vandalism.

3 What can be done with the extension?

The new tag `<R>` allows R programs to be embedded into a Mediawiki page. It is used to display raw output, table and graphics. The additional `<Rform>` tag allows for limited interactivity and basically masks the HTML `<form>` tag.

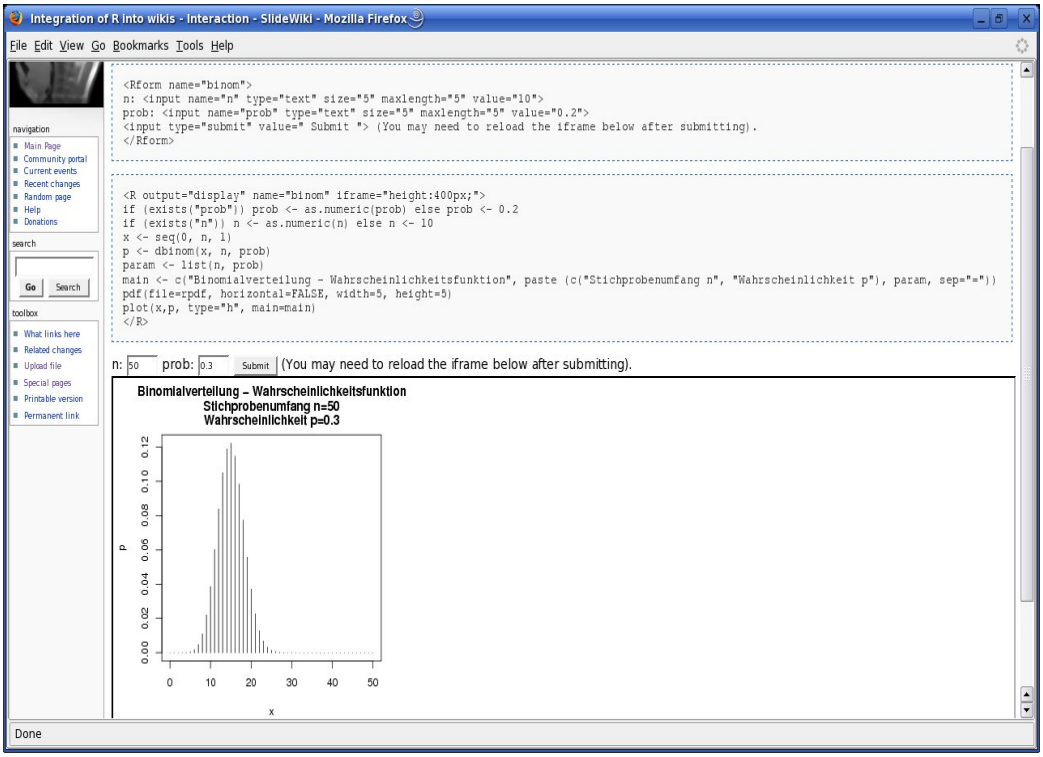

Figure 2: Interactive form for a graphics with the probability function of the Binomial distribution function computed by R in the WebPresentation wiki.

3.1 `<R>` tag

The opening `<R>` and the closing `</R>` tag embed an R program. Figure 1 shows how a simple R program is embedded in a Mediawiki page as well as the output of the program. The most important attributes from a users perspective are:

`display` determines which output of an R program is shown.

`echo` if set the R program is displayed before the program output.

`iframe` the program output is displayed in an iframe rather then a directly embedded.

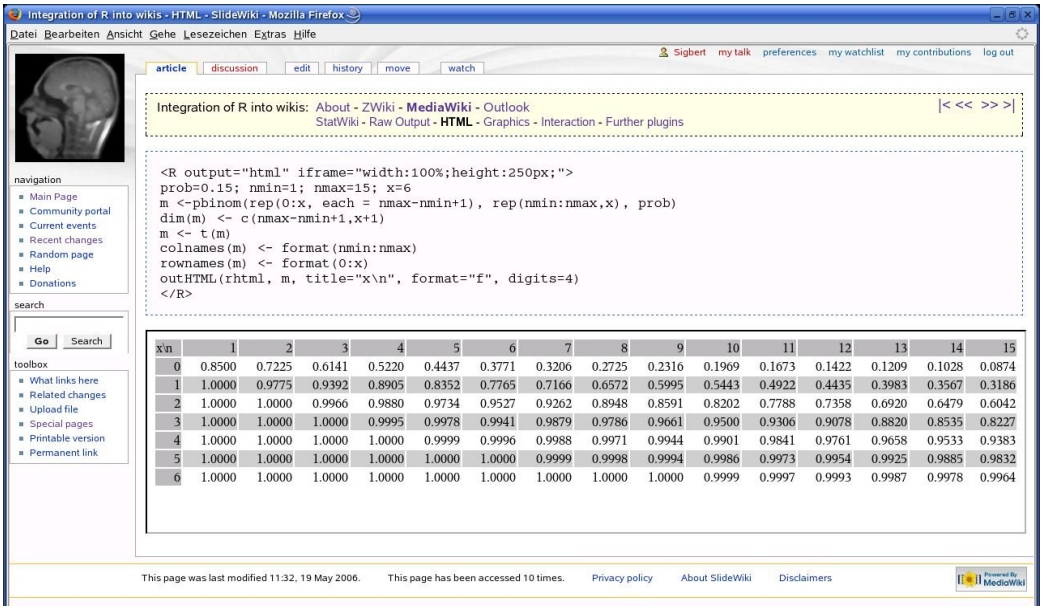

Figure 3: HTML table of the Binomial distribution function computed by R in the WebPresentation wiki.

`workspace` loads and saves a specific workspace at the program start and end.

The attribute `display` can take one of three values `text` (default), `html` or `display`. The first value should be used for the raw output from the R program, like in Figure 1. Figure 3 shows HTML output (`display="html"`) generated via the `outHTML` function. The `outHTML` function was developed to display nice tables generated from matrices. Finally Figure 4 shows an example of graphic output generated via the `pdf` output device of R.

For more detailled examples see in our WebPresentation wiki the presentation *Use of wikis in teaching - R plugin for Mediawiki*.

3.2 `<Rform>` tag

In StatWiki, (see below) we first made different tables for the distribution function for the Binomial distribution (p=0.05, 0.1, 0.15, ...), using several

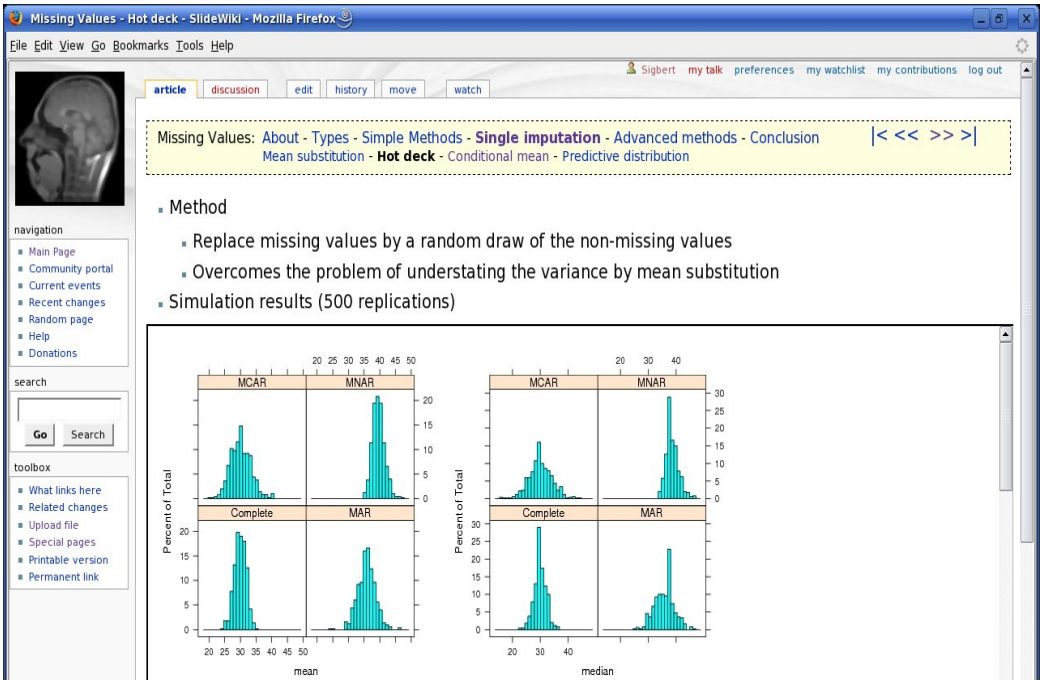

Figure 4: Graphical output from a R program in the WebPresentation wiki. 500 replications are done which show the effect of the "hot deck" imputation method on several parameters on a simulated data set.

R programs. Of course, we wanted to have an opportunity to provide parameters to the program interactively, in order to have just one table. The additional tag `<Rform>` allows to set parameters for a R program. The attributes `name` and `iframe` have also to be provided in the `<R>` tag. In Figure 2 the probability function of the Binomial distribution is shown for the different values of n and p. The parameters are entered in the R program as text values.

3.3 Mapping a user program to R

R program header and footer routines are added for the user. The header, depending on the options set in the `<R>` tag, consists of loading a workspace, sourcing `StatWiki.r` which, for example, contains the definition of `outHTML`, setting some variables (e.g. `rpdf`, `rhtml`, ...) and finally setting variables

```
# header
rfiles<-"/srv/.../Rfiles"
rpdf <- "/srv/.../Rfiles/76...c7.pdf"
n<-"50"     # these two lines appear only
prob<-"0.3" # if parameters are submitted
# user program
if (exists("prob"))
  prob <- as.numeric(prob) else prob <- 0.2
if (exists("n"))
  n <- as.numeric(n) else n <- 10
x <- seq(0, n, 1)
p <- dbinom(x, n, prob)
main <- "some headline ..."
pdf(rpdf)
plot(x,p, type="h", main=main)
# footer
q()
```

Figure 5: The R program which generates Figure 2 if the value 50 for `n` and 0.3 for `prob` are entered in the form and submitted.

from the `<Rform>` tag. The footer, also depending on the options set in the `<R>` tag, consists of saving the workspace and calling the `q()` routine; see the program in Figure 5.

Note that the extension is intended to embed short R programs. The standard initialisation file of PHP limits the execution time of an embedded PHP script to 30 seconds.

4 Problems

The use of the extension poses some standard problems with web-based applications.

4.1 Parallel access

In a web-based environment we have to cope with parallel access to the web pages. Everyone can edit a wiki page and therefore modify, submit and execute the embedded R program with saving the wiki page.

If a wiki page is edited by two different users then Mediawiki informs the user which saves last, that a modified version of the page has been saved before. The user has to decide if he wants to overwrite the modified page or not. However, to diminish the *last-one-wins* problem in Mediawiki wiki users can edit only a part of a page rather than the whole page.

A similar problem appears if two readers run the same interactive example. Assume, user 1 hits the *Submit* button and little bit later user 2, both using different parameters. If user 2 has a faster connection to the web server such that his program is executed at the web server first then it may happen that user 2 just sees the output of user 1.

The aim of the extension is to provide graphics and tables and use interaction where it helps to reduce the number of graphics and tables.

4.2 Security issues

Whenever you allow external users to send and execute programs on your web server it creates a potential security leak. Therefore the R extension has three different security levels:

0 The PHP extension and the R programs run under the same user as the web server and therefore the R program could read and overwrite the PHP scripts which are required to run Mediawiki.

1 (default security level). The R extension forbids the use of a set of R commands, based on the list of the R-php (Mineo and Pontillo, 2006) project. The one exception being `pdf` which is necessary to generate graphics.

2 The R extension runs the R programs under a separate user which will have no access to the PHP scripts which are required to run Mediawiki.

Please note: Anyone who installs the R extension should be aware that it is a potential security leak!

5 Applications

Our own installations and a some wikis which use our extension can be found on the internet; see the web links at the last section *Example wikis*.

WebPresentation - combines the idea of using slides (see the slide extension by Tels) with the R code for teaching and conference presentations; we used it for our presentation at the useR! 2006 conference in Vienna, Austria (see *Integration of R into wikis*).

StatWiki - is our dictionary wiki of statistical terms (in German). We use the extension within this to generate tables and graphics.

TeachWiki - is our wiki for final theses from statistics students (in German). The extension is used to generate graphics. This encourages students to use modern technology and allows us to transfer noteworthy theses to wikibooks on statistics.

StatClass - is a wiki by M. Cozowic for statistics exercises at university (in German). R is used to make computations etc.

WikiTraba - is a wiki about free software (in Spanish), which lists the extension under *Introducir programas de R* and shows results from it.

From our mailing list (`r-sig-mediawiki`) we have been able to find out that some intranet wikis have installed the extension:

- T.E. Cason combines the revision control and *real-time* publishing in a wiki with the computational possibilities of R in his work to report and to produce data snapshot analysis in clinical trials.
- The same potential as above is used by A. Brown in a different context (the tracking of bugs, regression testing and automated performance testing).

In both cases they moved from a situation (LaTeX+R+Sweave/Excel+Perl) where only one user had the data and access to the analysis procedure and results, to a wiki where everybody in the company had access to it.

5.1 Similar approaches for R

Since the main aim of the extension is to provide pictures, tables and occasionally R output to support text written in the Mediawiki the most similar is Sweave by Leisch (2002) and odfWeave by Kuhn and Weaston (2007). In both cases reports can be generated for HTML and LaTeX by Sweave and for OpenOffice by odfWeave.

There are two differences between Sweave and our extension:

- Sweave extracts the R code chunks from the Sweave file and inserts the result of the R code chunks into a new LaTeX file. The Mediawiki extension mechanism already provides us with the R code chunks which then only need to be evaluated.
- The extension allows for simple interaction whereas the nature of LaTeX does not require this from Sweave.

On one hand we have programs like Sweave which runs inside R, on the other hand we have packages like R-php (Mineo and Pontillo 2006) which embed R programs, like our extension, into PHP which itself is embbeded in HTML pages. In same manner as R-php we evaluate the R code chunks with the batch version of R. Although the Mediawiki is written itself in PHP, we do have not PHP as a basis but instead a text which follows the wiki syntax with embedded R programs.

Thus we place our extension somewhere between Sweave and R-php.

6 Improvements

Several improvements, modifications or extensions has been asked by users

Sweave functionality

At present the R extension only allows for the R program to be displayed at the start, however it would be more desirable if, as in Sweave, we could also display sequences of R input, R output, R input, R output, ...

Offline version

Wikis always require open internet connections and cannot be used by offline computers. We could provide either a WAMPP package or a virtual PC with all the required tools and some example wikis.

Extension to other wikis

Moving as much code as possible into R would open the possibility of extending the extension to other PHP based wikis. This would be particularly interesting for the DokuWiki, the basis of the R Wiki.

7 Acknowledgements

We would like to thank A. Browne, T.E. Cason, M. Cassin, M. Cozowicz, E. Neuwirth, L. Udvarhelyi and the unknown referees for their support in developing and improving the extension as well as for their support for writing this article.

References and web links

Kuhn, M., Weaston, S. (2007). odfWeave: Sweave processing of Open Document Format (ODF) files. CRAN package,
http://cran.r-project.org/src/contrib/Descriptions/odfWeave.html

Leisch, F. (2007), R behind the scenes: Using S the (un)usual way, Proceedings of the ISI 2007, Lisbon, forthcoming

Leisch, F. (2002), Sweave: Dynamic generation of statistical reports. In: W. Hrdle and B. Rnz, Hrgs., CompStat 2002 - Proceedings in Computational Statistics, p. 575-580. Physica Verlag, Heidelberg,

http://www.ci.tuwien.ac.at/ leisch/Sweave

Ludwig-Mayerhofer, L. (2004), ILMES - Internet-Lexikon der Methoden der empirischen Sozialforschung, Virtuelle Fachbibliothek Psychologie, Saarlndische Universitts- und Landesbibliothek,
http://psydok.sulb.uni-saarland.de/volltexte/2004/260

Mineo, A., Pontillo, A. (2006), Using R via PHP: R-php, The R-User conference 2006, Book of abstracts, p. 121,
http://www.r-project.org/user-2006/Abstracts/Mineo+Pontillo.pdf

R extension

Download:
• http://mars.wiwi.hu-berlin.de/mediawiki/sk/index.php/R_extension_for_Mediawiki
Mailinglist:
• https://stat.ethz.ch/mailman/listinfo/r-sig-mediawiki

Mediawiki

Mediawiki:
• http://www.mediawiki.org
Wikipedia:
• http://www.wikipedia.org
Slide extension:
• http://bloodgate.com/wiki/Wiki-Presentations

Example wikis

WebPresentation:
• http://mars.wiwi.hu-berlin.de/mediawiki/slides
StatWiki:
• http://statwiki.wiwi.hu-berlin.de
TeachWiki:

• http://teachwiki.wiwi.hu-berlin.de
TrabaWiki:
• http://www.traba.org
StatClass:
• http://eisber.net/StatWiki

SFB 649 Discussion Paper Series 2007

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Trade Liberalisation, Process and Product Innovation, and Relative Skill Demand" by Sebastian Braun, January 2007.
002 "Robust Risk Management. Accounting for Nonstationarity and Heavy Tails" by Ying Chen and Vladimir Spokoiny, January 2007.
003 "Explaining Asset Prices with External Habits and Wage Rigidities in a DSGE Model." by Harald Uhlig, January 2007.
004 "Volatility and Causality in Asia Pacific Financial Markets" by Enzo Weber, January 2007.
005 "Quantile Sieve Estimates For Time Series" by Jürgen Franke, Jean-Pierre Stockis and Joseph Tadjuidje, February 2007.
006 "Real Origins of the Great Depression: Monopolistic Competition, Union Power, and the American Business Cycle in the 1920s" by Monique Ebell and Albrecht Ritschl, February 2007.
007 "Rules, Discretion or Reputation? Monetary Policies and the Efficiency of Financial Markets in Germany, 14th to 16th Centuries" by Oliver Volckart, February 2007.
008 "Sectoral Transformation, Turbulence, and Labour Market Dynamics in Germany" by Ronald Bachmann and Michael C. Burda, February 2007.
009 "Union Wage Compression in a Right-to-Manage Model" by Thorsten Vogel, February 2007.
010 "On σ−additive robust representation of convex risk measures for unbounded financial positions in the presence of uncertainty about the market model" by Volker Krätschmer, March 2007.
011 "Media Coverage and Macroeconomic Information Processing" by Alexandra Niessen, March 2007.
012 "Are Correlations Constant Over Time? Application of the CC-TRIG$_t$-test to Return Series from Different Asset Classes." by Matthias Fischer, March 2007.
013 "Uncertain Paternity, Mating Market Failure, and the Institution of Marriage" by Dirk Bethmann and Michael Kvasnicka, March 2007.
014 "What Happened to the Transatlantic Capital Market Relations?" by Enzo Weber, March 2007.
015 "Who Leads Financial Markets?" by Enzo Weber, April 2007.
016 "Fiscal Policy Rules in Practice" by Andreas Thams, April 2007.
017 "Empirical Pricing Kernels and Investor Preferences" by Kai Detlefsen, Wolfgang Härdle and Rouslan Moro, April 2007.
018 "Simultaneous Causality in International Trade" by Enzo Weber, April 2007.
019 "Regional and Outward Economic Integration in South-East Asia" by Enzo Weber, April 2007.
020 "Computational Statistics and Data Visualization" by Antony Unwin, Chun-houh Chen and Wolfgang Härdle, April 2007.
021 "Ideology Without Ideologists" by Lydia Mechtenberg, April 2007.
022 "A Generalized ARFIMA Process with Markov-Switching Fractional Differencing Parameter" by Wen-Jen Tsay and Wolfgang Härdle, April 2007.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

023 "Time Series Modelling with Semiparametric Factor Dynamics" by Szymon Borak, Wolfgang Härdle, Enno Mammen and Byeong U. Park, April 2007.
024 "From Animal Baits to Investors' Preference: Estimating and Demixing of the Weight Function in Semiparametric Models for Biased Samples" by Ya'acov Ritov and Wolfgang Härdle, May 2007.
025 "Statistics of Risk Aversion" by Enzo Giacomini and Wolfgang Härdle, May 2007.
026 "Robust Optimal Control for a Consumption-Investment Problem" by Alexander Schied, May 2007.
027 "Long Memory Persistence in the Factor of Implied Volatility Dynamics" by Wolfgang Härdle and Julius Mungo, May 2007.
028 "Macroeconomic Policy in a Heterogeneous Monetary Union" by Oliver Grimm and Stefan Ried, May 2007.
029 "Comparison of Panel Cointegration Tests" by Deniz Dilan Karaman Örsal, May 2007.
030 "Robust Maximization of Consumption with Logarithmic Utility" by Daniel Hernández-Hernández and Alexander Schied, May 2007.
031 "Using Wiki to Build an E-learning System in Statistics in Arabic Language" by Taleb Ahmad, Wolfgang Härdle and Sigbert Klinke, May 2007.
032 "Visualization of Competitive Market Structure by Means of Choice Data" by Werner Kunz, May 2007.
033 "Does International Outsourcing Depress Union Wages? by Sebastian Braun and Juliane Scheffel, May 2007.
034 "A Note on the Effect of Outsourcing on Union Wages" by Sebastian Braun and Juliane Scheffel, May 2007.
035 "Estimating Probabilities of Default With Support Vector Machines" by Wolfgang Härdle, Rouslan Moro and Dorothea Schäfer, June 2007.
036 "Yxilon – A Client/Server Based Statistical Environment" by Wolfgang Härdle, Sigbert Klinke and Uwe Ziegenhagen, June 2007.
037 "Calibrating CAT Bonds for Mexican Earthquakes" by Wolfgang Härdle and Brenda López Cabrera, June 2007.
038 "Economic Integration and the Foreign Exchange" by Enzo Weber, June 2007.
039 "Tracking Down the Business Cycle: A Dynamic Factor Model For Germany 1820-1913" by Samad Sarferaz and Martin Uebele, June 2007.
040 "Optimal Policy Under Model Uncertainty: A Structural-Bayesian Estimation Approach" by Alexander Kriwoluzky and Christian Stoltenberg, July 2007.
041 "QuantNet – A Database-Driven Online Repository of Scientific Information" by Anton Andriyashin and Wolfgang Härdle, July 2007.
042 "Exchange Rate Uncertainty and Trade Growth - A Comparison of Linear and Nonlinear (Forecasting) Models" by Helmut Herwartz and Henning Weber, July 2007.
043 "How do Rating Agencies Score in Predicting Firm Performance" by Gunter Löffler and Peter N. Posch, August 2007.

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

044 "Ein Vergleich des binären Logit-Modells mit künstlichen neuronalen Netzen zur Insolvenzprognose anhand relativer Bilanzkennzahlen" by Ronald Franken, August 2007.
045 "Promotion Tournaments and Individual Performance Pay" by Anja Schöttner and Veikko Thiele, August 2007.
046 "Estimation with the Nested Logit Model: Specifications and Software Particularities" by Nadja Silberhorn, Yasemin Boztuğ and Lutz Hildebrandt, August 2007.
047 "Risiken infolge von Technologie-Outsourcing?" by Michael Stephan, August 2007.
048 "Sensitivities for Bermudan Options by Regression Methods" by Denis Belomestny, Grigori Milstein and John Schoenmakers, August 2007.
049 "Occupational Choice and the Spirit of Capitalism" by Matthias Doepke and Fabrizio Zilibotti, August 2007.
050 "On the Utility of E-Learning in Statistics" by Wolfgang Härdle, Sigbert Klinke and Uwe Ziegenhagen, August 2007.
051 "Mergers & Acquisitions and Innovation Performance in the Telecommunications Equipment Industry" by Tseveen Gantumur and Andreas Stephan, August 2007.
052 "Capturing Common Components in High-Frequency Financial Time Series: A Multivariate Stochastic Multiplicative Error Model" by Nikolaus Hautsch, September 2007.
053 "World War II, Missing Men, and Out-of-wedlock Childbearing" by Michael Kvasnicka and Dirk Bethmann, September 2007.
054 "The Drivers and Implications of Business Divestiture – An Application and Extension of Prior Findings" by Carolin Decker, September 2007.
055 "Why Managers Hold Shares of Their Firms: An Empirical Analysis" by Ulf von Lilienfeld-Toal and Stefan Ruenzi, September 2007.
056 "Auswirkungen der IFRS-Umstellung auf die Risikoprämie von Unternehmensanleihen - Eine empirische Studie für Deutschland, Österreich und die Schweiz" by Kerstin Kiefer and Philipp Schorn, September 2007.
057 "Conditional Complexity of Compression for Authorship Attribution" by Mikhail B. Malyutov, Chammi I. Wickramasinghe and Sufeng Li, September 2007.
058 "Total Work, Gender and Social Norms" by Michael Burda, Daniel S. Hamermesh and Philippe Weil, September 2007.
059 "Long-Term Orientation in Family and Non-Family Firms: a Bayesian Analysis" by Jörn Hendrich Block and Andreas Thams, October 2007
060 "Kombinierte Liquiditäts- und Solvenzkennzahlen und ein darauf basierendes Insolvenzprognosemodell für deutsche GmbHs" by Volodymyr Perederiy, October 2007
061 "Embedding R in the Mediawiki" by Sigbert Klinke and Olga Zlatkin-Troitschanskaia, October 2007

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".