

Electronic books for experts and users

Zdeněk Hlávka, Sigbert Klinke, Rodrigo Witzel

Institut für Statistik und Ökonometrie
Humboldt-Universität zu Berlin
email: hlavka@wiwi.hu-berlin.de
www.md-stat.com



Outline of the talk

Electronic book:

1. formats/what is e-book?
2. production
 - MD*Book
 - modes and conditions
 - MD*Book online
3. new e-book format
4. distribution of e-books

Electronic book complements the printed version of the book.

Usually one wants to produce the following formats:

postscript for printed version of the book,

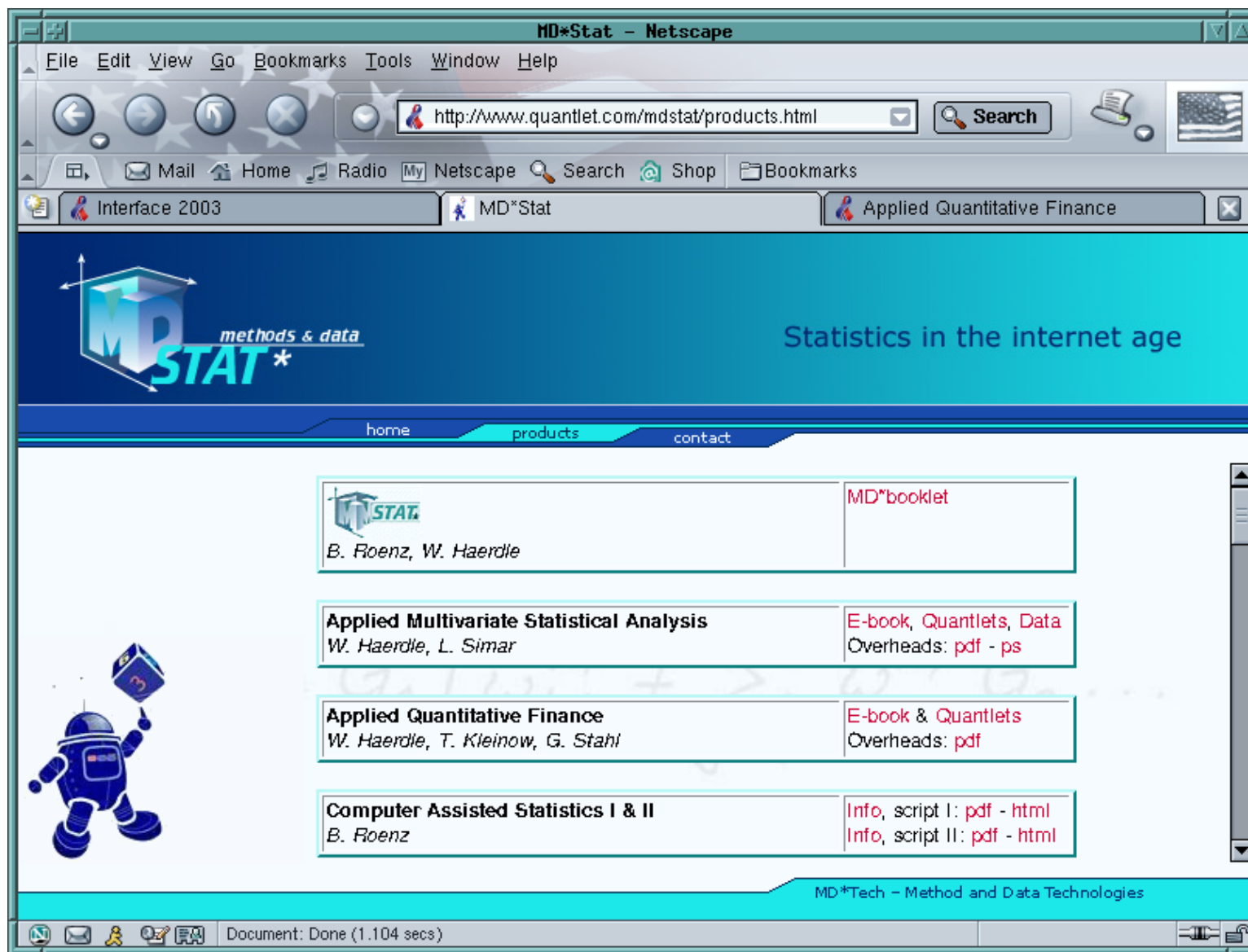
pdf for downloadable e-book,

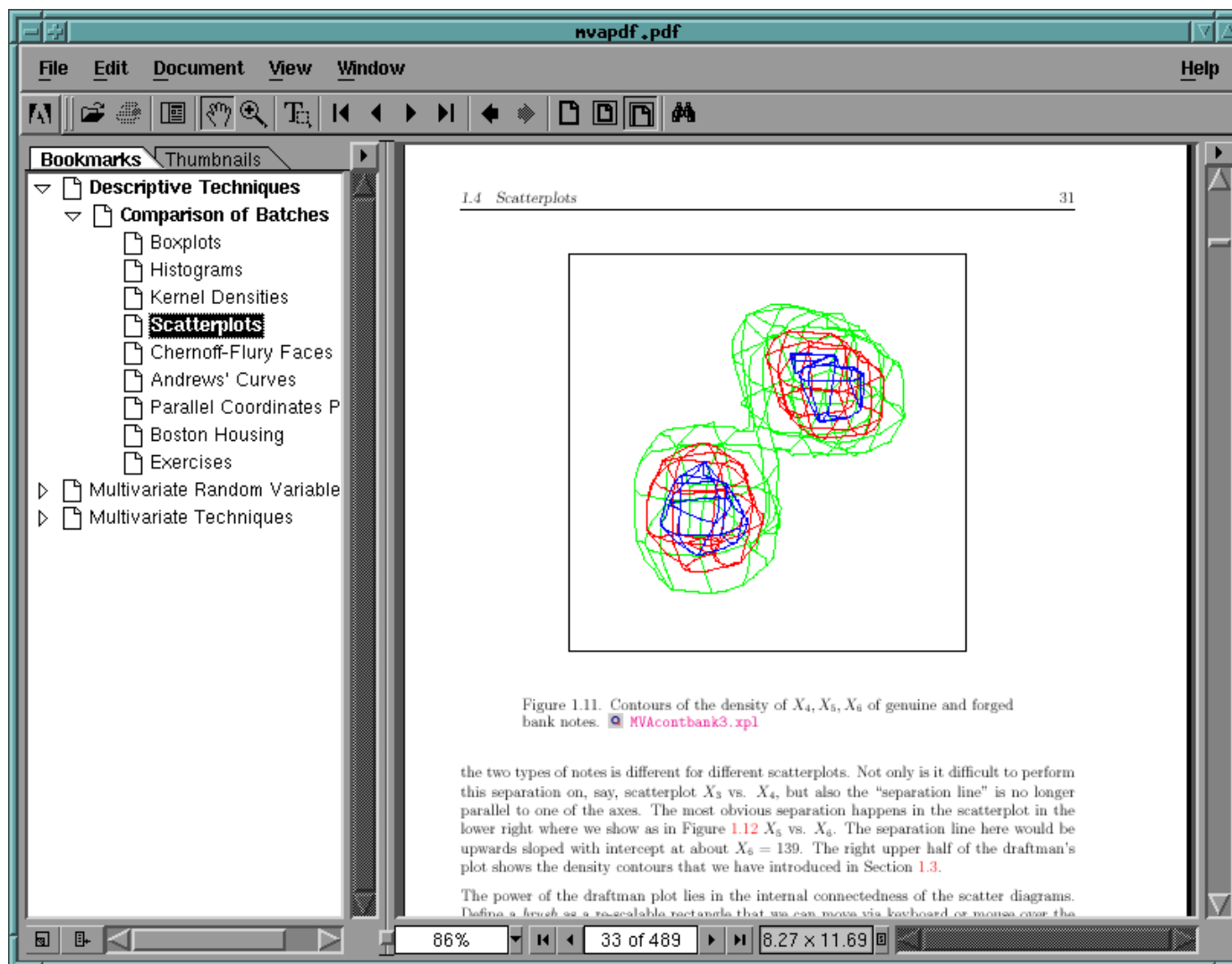
html for online e-book.

Add-ons:

- another text containing related information (e.g. slides or exercises)
- interactive features making the text more attractive

The following examples are available at www.md-stat.com.



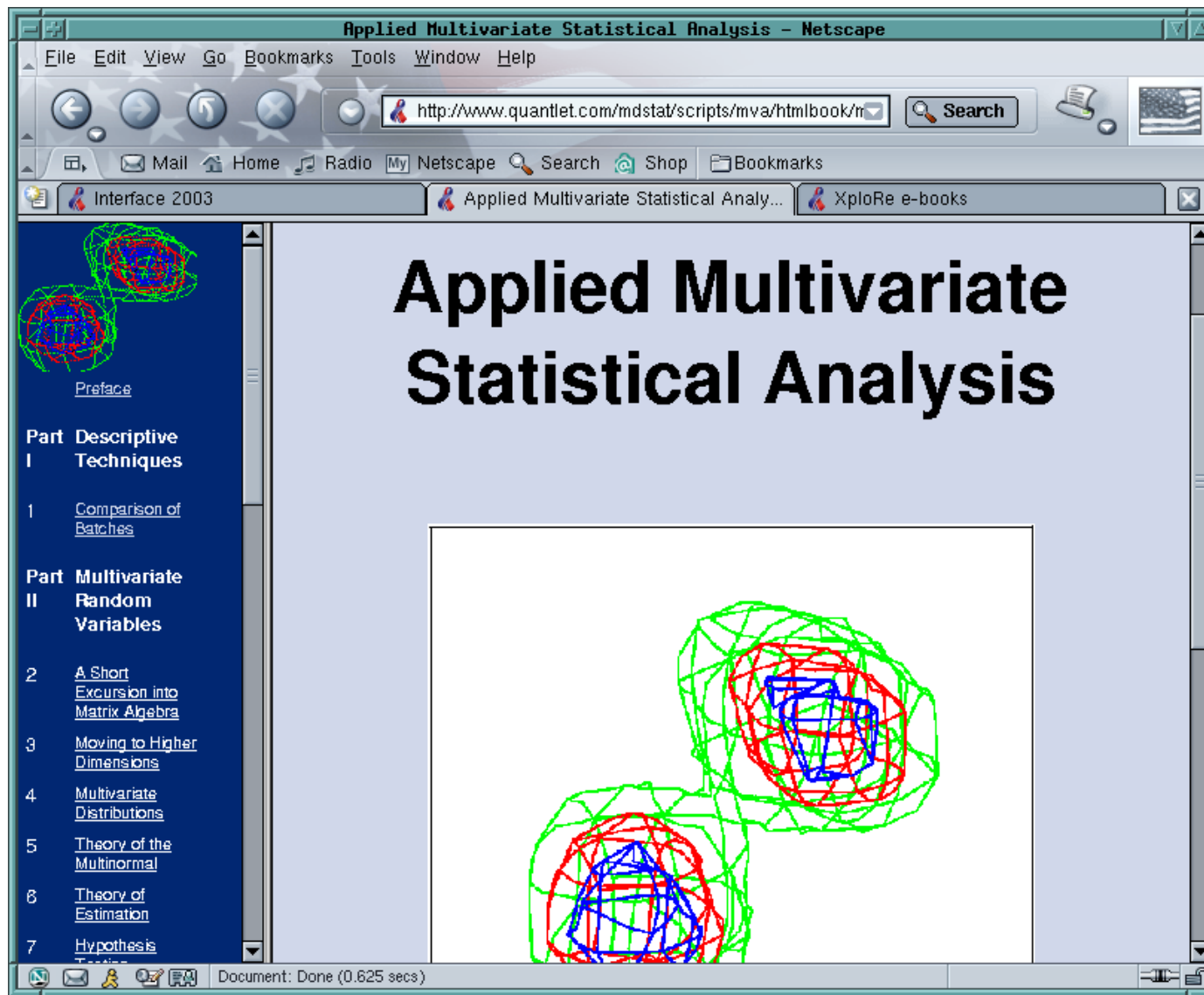


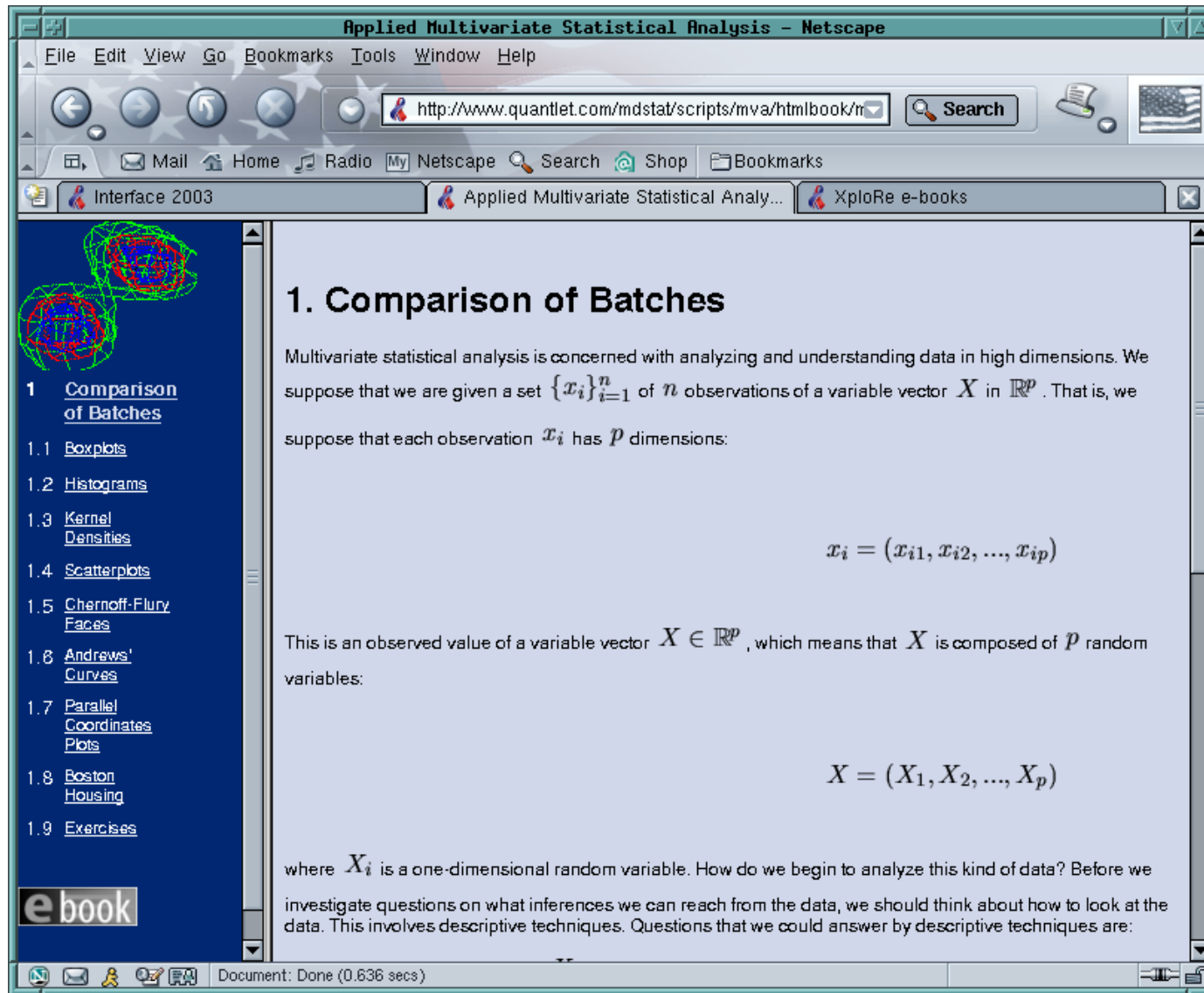
PDF format

- The text layout is identical to the printed version of the book.
- It can be downloaded via internet.
- In addition, it contains hyperlinks and navigation menu.

It inherits a lot of (too many?) properties of the printed book.

The structure of the document is affected e.g. by pagesize which doesn't make much sense if you read the document on monitor.





The screenshot shows a Netscape browser window with the title "Applied Multivariate Statistical Analysis - Netscape". The address bar contains the URL "http://www.quantlet.com/mdstat/scripts/mva/htmlbook/m". The browser interface includes a menu bar (File, Edit, View, Go, Bookmarks, Tools, Window, Help), a toolbar with navigation buttons, and a search box. The page content is displayed in a two-column layout. The left column is a dark blue sidebar with a table of contents and an "ebook" logo. The right column contains the main text of the chapter, including a definition of multivariate statistical analysis, a mathematical expression for a variable vector x_i , and a definition of a variable vector X .

1. Comparison of Batches

Multivariate statistical analysis is concerned with analyzing and understanding data in high dimensions. We suppose that we are given a set $\{x_i\}_{i=1}^n$ of n observations of a variable vector X in \mathbb{R}^p . That is, we suppose that each observation x_i has p dimensions:

$$x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$$

This is an observed value of a variable vector $X \in \mathbb{R}^p$, which means that X is composed of p random variables:

$$X = (X_1, X_2, \dots, X_p)$$

where X_i is a one-dimensional random variable. How do we begin to analyze this kind of data? Before we investigate questions on what inferences we can reach from the data, we should think about how to look at the data. This involves descriptive techniques. Questions that we could answer by descriptive techniques are:

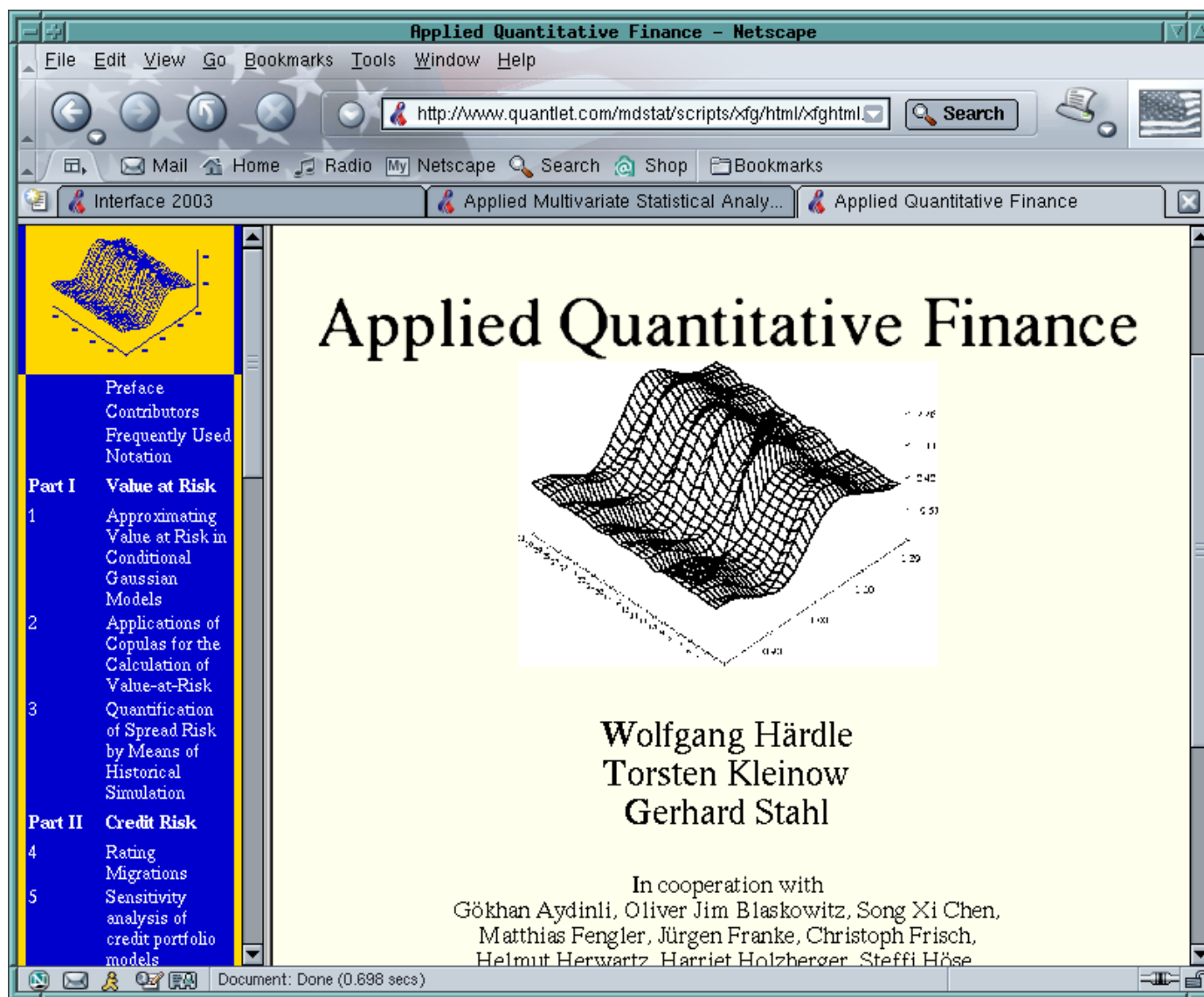
1. Comparison of Batches

- 1.1 [Boxplots](#)
- 1.2 [Histograms](#)
- 1.3 [Kernel Densities](#)
- 1.4 [Scatterplots](#)
- 1.5 [Chernoff-Flury Faces](#)
- 1.6 [Andrews' Curves](#)
- 1.7 [Parallel Coordinates Plots](#)
- 1.8 [Boston Housing](#)
- 1.9 [Exercises](#)

Document: Done (0.636 secs)

HTML format

- The formulas don't look as good as in PDF.
- The producer doesn't have full control the layout of the page.
- E-book structure follows the structure of the text (no paper size).
- It is possible to include “interactive components”
(as in the new e-book format presented in Part 3 of the talk).
- The look of the page can be controlled with CSS files.



It is hard to decide which e-book format is the best one, but there is clearly a need for a system which allows more output formats:

1. postscript
2. pdf
3. html
4. other formats

Definition: Electronic book is a file which can be used as a source for the above formats.

MD*Book is a set of perl scripts which use existing tools such as pdflatex and latex2html to create the desired versions of the document (it is unfeasible to start programming such a system from a scratch).

MD*Book manages many output formats automatically.

It is based on LaTeX and on free translation programs such as pdflatex and latex2html.

In its simplest form, it has only command-line interface (Linux) which offers many possibilities of influencing the behavior of the system.

```
mdbook2 -format inputfile.sk
```

Input

- TeX files containing the text (in LaTeX format, sectionwise),
- .sk file which says how to put the TeX files together.

MD*Book command on the commandline

```
mdbook2 -ps -html inputfile.sk
```

~> generate the ps and html version of the e-book using “inputfile.sk”.

Simplified example of .sk file:

```
USEPACKAGE      html
BEGIN ps pdf
USEPACKAGE      hyperref
END
FINALPS         psbook/mvatex.ps
FINALPDF        pdfbook/mvapdf.pdf
FINALHTML       htmlbook/mvahtml.html
```

```
FORMAT          scrbook 12pt a4paper headsepline
START           title
PART
TITLE           Descriptive Techniques
CHAPTER         c1
TITLE           Comparison of Batches
LABEL          ch01
SECTION         c1boxin
TITLE           Boxplots
LABEL          c1boxin
SECTION         c1histo
TITLE           Histograms
INDEX           histograms
BIBLIOGRAPHY   mm
PRINTINDEX
END
```

How does it work?

After great simplification:

1. use the .sk file to create ordinary TeX file
2. run:
 - latex and dvips to obtain the postscript version
 - pdflatex to obtain the pdf version
 - latex2html to obtain the html version

MD*Book is necessary because certain LaTeX commands have to be defined differently for pdf, ps, and html versions.

The .sk file contains information on the structure of the document which is used to generate e.g. the navigation in the resulting html document.

Further keywords allow better management of input and output files.

MODE

Using the keyword `MODE` in the `.sk` file allows to have more documents with the same structure containing different text.

We use this feature e.g. for our lectures: the script and the slides have always the same structure.

The script and the slides have common `.sk` file (common structure).

```
MODE slide
```

switches to the slide mode: the “slide” TeX files are used.

The .sk file contains the following construction

```
BEGIN      slide
MODE       slide
END
```

We switch to the slide mode if the call to MD*Book looks like this

```
mdbook2 -ps -pdf inputfile.sk slide
```

The following part of the .sk file

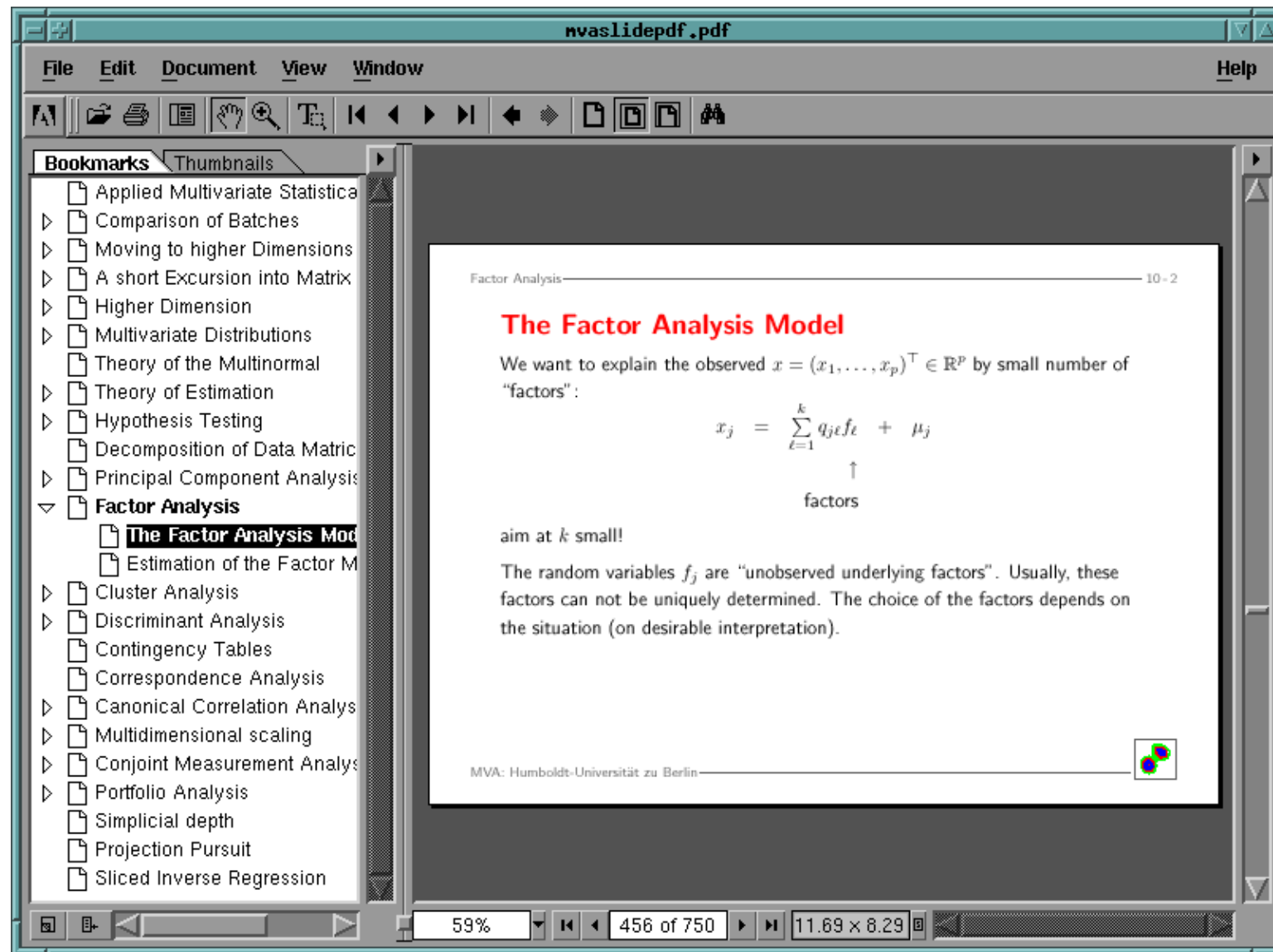
```
SECTION      c1boxin
TITLE        Boxplots
```

is then understood as

```
\section{Boxplots}
\input{c1boxinslide}
```

instead of

```
\section{Boxplots}
\input{c1boxin}
```



The screenshot shows a PDF viewer window titled "nvaslidepdf .pdf". The menu bar includes "File", "Edit", "Document", "View", "Window", and "Help". The toolbar contains various navigation and editing icons. On the left, a "Bookmarks" panel is visible, listing a hierarchy of topics. The main content area displays a slide titled "The Factor Analysis Model". The slide text includes a definition of the observed vector $x = (x_1, \dots, x_p)^T \in \mathbb{R}^p$ and a mathematical equation for x_j as a sum of products of loadings and factors, plus an error term. It also states the goal of having a small number of factors and notes that the factors are unobserved and not uniquely determined. The footer of the slide mentions "MVA: Humboldt-Universität zu Berlin" and includes a small logo.

Factor Analysis — 10-2

The Factor Analysis Model

We want to explain the observed $x = (x_1, \dots, x_p)^T \in \mathbb{R}^p$ by small number of "factors":

$$x_j = \sum_{\ell=1}^k q_{j\ell} f_{\ell} + \mu_j$$

↑
factors

aim at k small!

The random variables f_j are "unobserved underlying factors". Usually, these factors can not be uniquely determined. The choice of the factors depends on the situation (on desirable interpretation).

MVA: Humboldt-Universität zu Berlin

The above presented tools are very powerful, but some simplification is needed.

tex2sk is a script which creates the .sk file automatically so that normal user doesn't have to know anything about the .sk file.

Thus all above output formats can be obtained without knowing that the .sk exists.

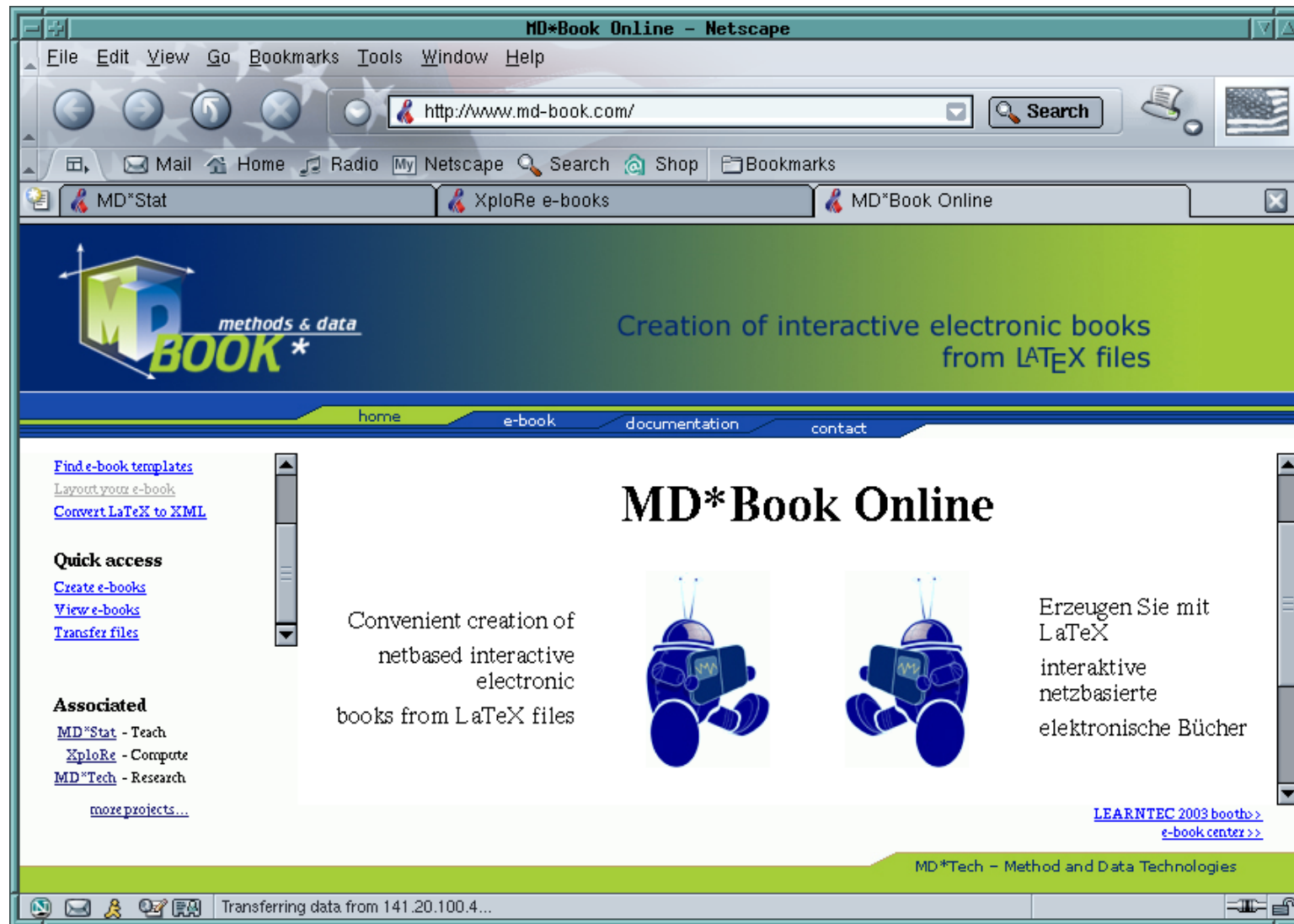
Problems

The biggest problem is the dependency on the pdflatex and latex2html commands. These scripts keep changing and mdbook2 is thus rather fragile.

We can not guarantee that mdbook2 works well with any version of latex, pdflatex, and latex2html.

There are many potential users who want to use other operating systems.

We offer MD*Book online service at www.md-book.com which allows anybody to compile the files on our server.



MD*Book online

www.md-book.com

In ideal case:

1. create login
2. create e-book project
3. transfer your TeX files to our server
4. create desired output format by clicking on the appropriate button
5. download the resulting files

In praxis, the TeX files usually need some tuning.

MD*Booklet

New e-book formats are possible and desirable.

E-stat project is based on XML.

- It is still in the early stages of development.
- Some information can be found at www.emilea.de.

MD*Booklet is based on HTML and Jscript.

- It offers more ways of navigation through the document than HTML.
- This format can be generated from a LaTeX file using MD*Book. The look of the resulting e-book can be affected by many tuning parameters.
- Many examples can be found at www.md-stat.com.



Microsoft Internet Explorer

lecture contents lecture 1.1 lecture 1.2 lecture 1.3 lecture 4.2 lecture 6.3 enhanced 6.3

6.3 Binomial Distribution

A Binomial distribution is derived from a **random experiment** in which we either obtain **event** A with constant **probability** p , or the **complementary event** \bar{A} with probability $1-p$.

Suppose this experiment is repeated n times.

A **discrete random variable** that contains the number of successes A after n repetitions of this experiment, has a Binomial distribution with parameters n and p . Its probability density function is:

$$f(x; n, p) = \begin{cases} \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} & \text{for } x = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

Denoted: $X \sim B(n; p)$

contents information explained enhanced enhanced interactive

Enhanced example for Binomial distribution

XploRe Introductory Course - Microsoft Internet Explorer

Course contents

Contents

1. Introduction to XploRe	1.1	The Desktop
2. XploRe and its Programming Language	1.2	The Input Window
3. Objects, Operators, Functions	1.3	The Output Window
4. Input and output of data	1.4	The Editor Window
5. Descriptive statistics	1.5	Running an example
6. Graphics	1.6	Auto Pilot Support System

contents

Finance Introductory Course - Microsoft Internet Explorer

FINANCE INTRODUCTORY COURSE

Matthias Fengler and Wolfgang Härdle

Course Contents Course 1.1 Course 1.2 **Course 2.1** Course 3.1

The owner of a **zerobond** buys this bond in t_0 paying B_0 and will receive an amount of B_T at time T . B_T is called the face value of the bond. B_T is non random.

The owner of a **coupon bond** will receive an amount of B_T at time T and additional payments at discrete time points t_0, \dots, t_n before T . The additional payments are called coupons.

Computing components

The resulting documents can include also computing/interactive components.

In MD*Book, this is currently solved by links from the documents to the computing/interactive examples which are included in separate HTML pages.

Such pages can be also automatically generated by MD*Book.

Applied Multivariate Statistical Analysis - Netscape

File Edit View Go Bookmarks Tools Window Help

http://www.quantlet.com/mdstat/codes/mva/MVAandcur2.html Search

Applied Multivariate Statistical Analy... XploRe e-books MD*Book Online

QUANTLETS

home execute edit help

MVAandcur2

Description: MVAandcur2 computes andrew's curves for the observations 96-105 of the Swiss bank notes data ("bank2.dat"). Here we changed the order of the variables

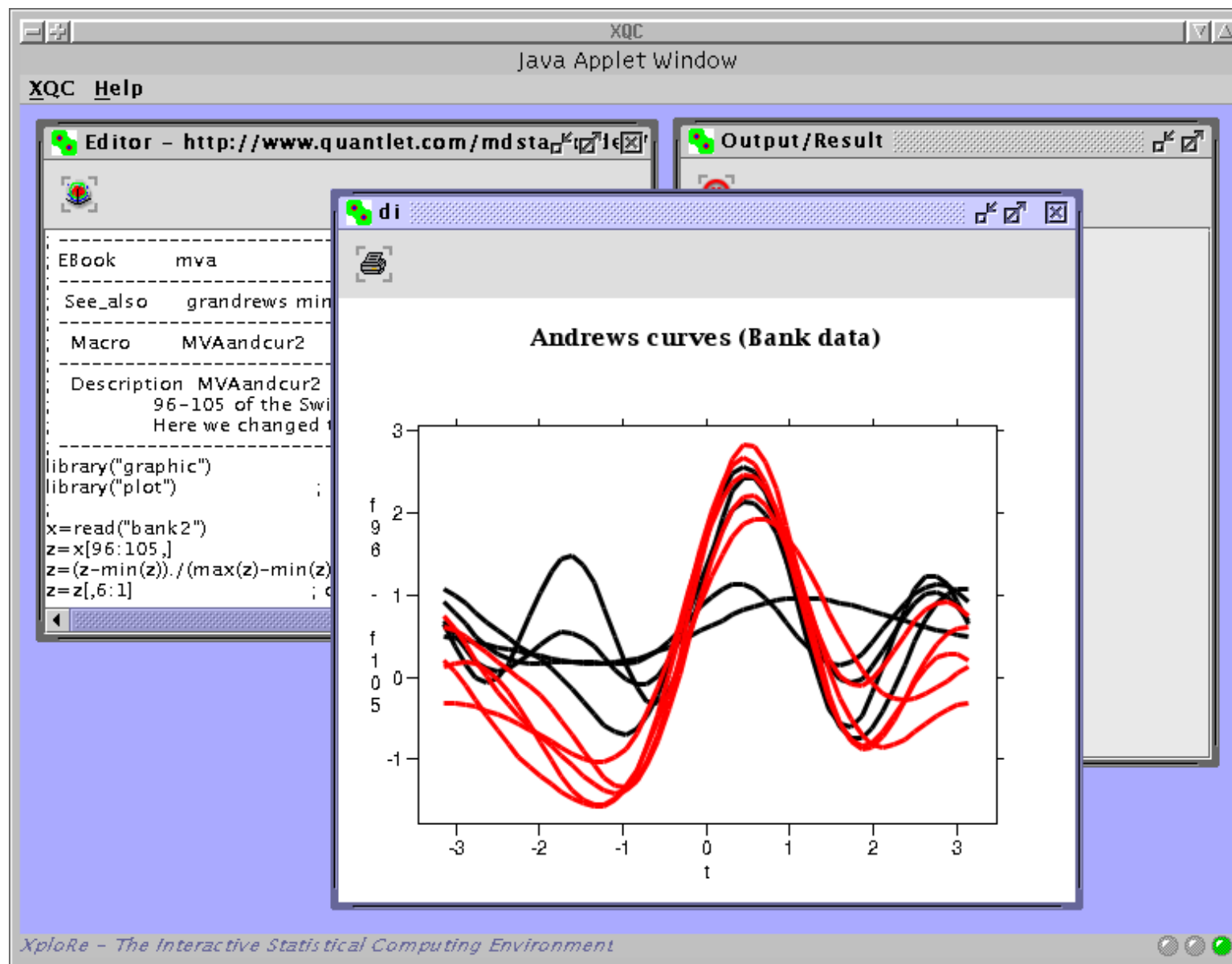
Download: [MVAandcur2.xpl](#)

Code:

```
library("graphic")           ; loads the graphic library
library("plot")             ; loads the plot library
x=read("bank2")             ; reads data
z=x[96:105,]                ; extracts observations 96-105
```

MD*Tech Method and Data Technologies

Document: Done (0.691 secs)



Distribution

On-line HTML, MD*Booklet.

Free download PDF

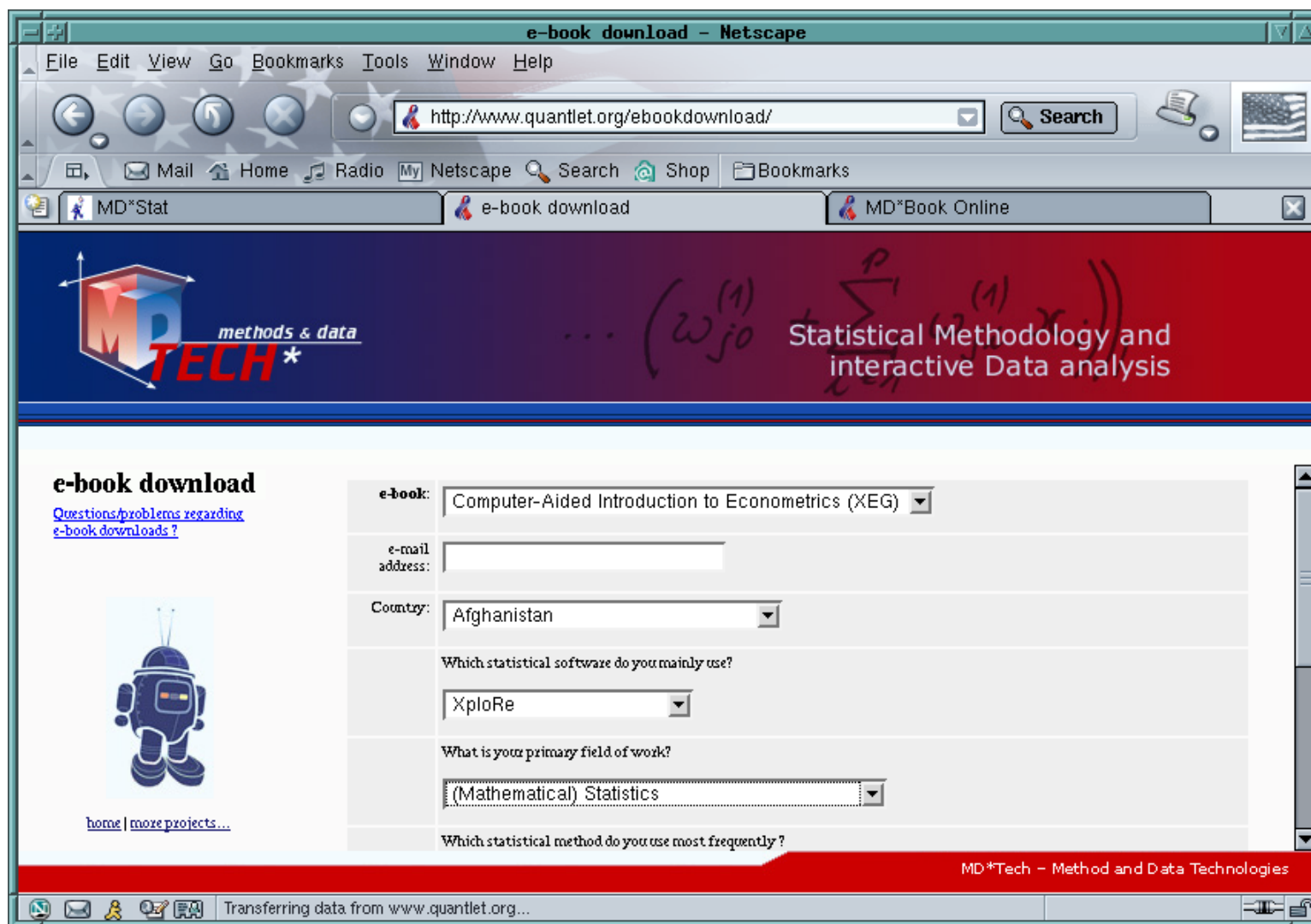
www.md-stat.com.

CD-ROM in the printed version of the book

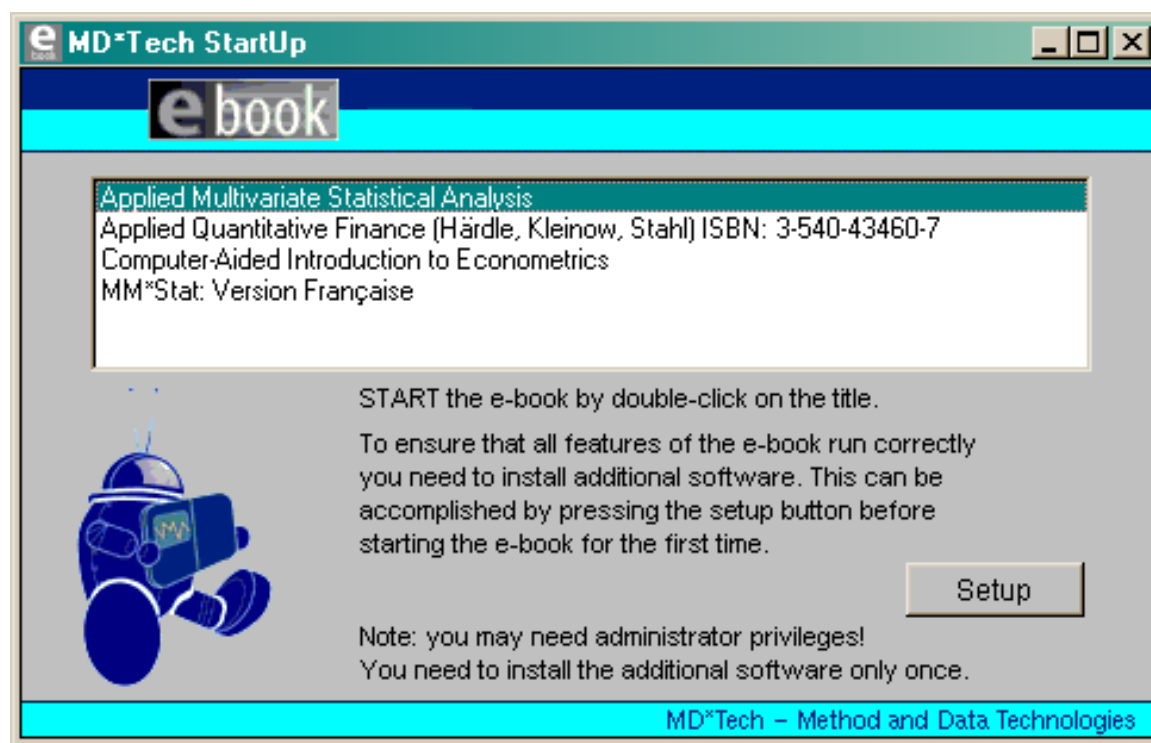
automatically generated with MD*Book (links are different from online version).

Download version for registered users,

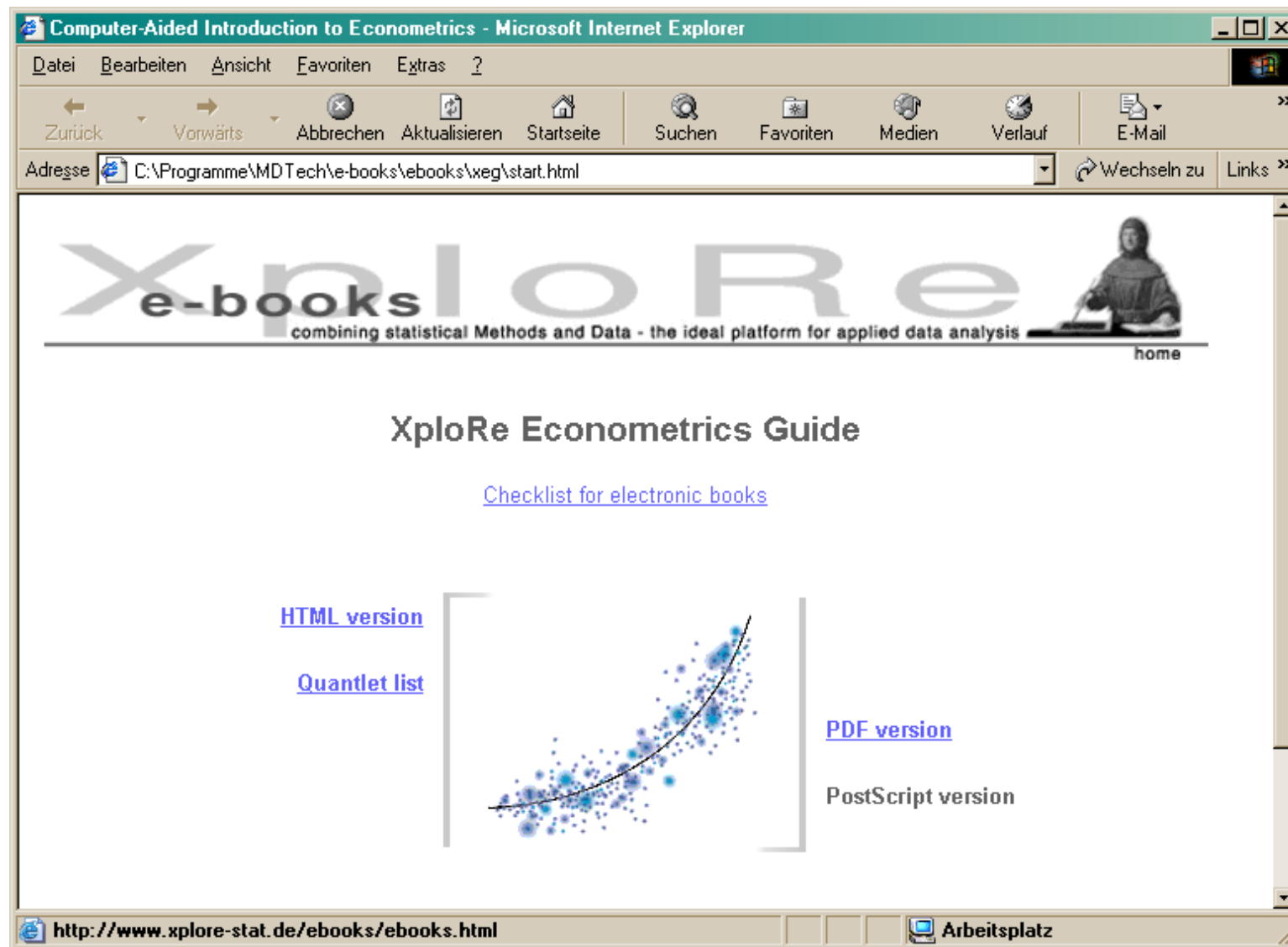
serial number necessary for download and installation is inserted in the printed version of the book.



After installing the downloaded executable file (30MB), it is possible to start the e-book.



In this case, also the XQS is started so that the included interactive and computing examples can be executed locally.



Conclusion



The publishing house Springer-Verlag will offer this technology in near future.

Check the following www pages for more information:

- www.md-stat.com,
- www.md-book.com.

It is not clear which of the current formats (such as Open eBook or Adobe eBook) will become the standard in the next decade, but MD*Book is certainly useful tool that can adapt and survive.