Multivariate Methods in Finance and Marketing
Lecture Notes

Prof. Dr. Joachim Grammig and M. Sc. Miriam Sperl
Department of Statistics, Econometrics and Empirical Economics
University of Tübingen

Winter Term 08/09

Learning Objectives

- This course introduces students to the **application of multivariate methods in finance and marketing** by means of various **case studies**, primarily chosen from market and empirical capital market research.

- This course teaches students how to **conduct empirical analyses** on their own: from the choice of the appropriate multivariate method over the correct data handling using the software package SAS to the competent interpretation and commentation on the results.

- This course reveals **manipulation possibilities** within the various methods and enables students to **question results critically**.

Course Organization

- **Lecture**:
  Tuesdays from 2:15 p.m. to 3:45 p.m.
  room E04 (Mohlstraße 36)

- **Practical Class:**
  two alternating groups covering exactly the same material
  Tuesdays from 4:15 p.m. to 5:45 p.m. at fortnightly intervals
  PC-Lab (Nauklerstraße 47)

- **Office Hours:**
  upon appointment with M. Sc. Miriam Sperl
  Department of Statistics, Econometrics and Empirical Economics
  room 316 (Mohlstraße 36)
  mail to miriam.sperl@uni-tuebingen.de

Course Information

- **Recommendations:**
  the course is recommended for 3rd year or more advanced Bachelor or Diploma students.

- **Prerequisites:**
  undergraduate mathematics and statistics.

- **ECTS or Credit Points:**
  7.5 ECTS for Bachelor or 5 LP for Diploma students can be obtained via successfully passing the final exam (written) at the end of this winter term or at the beginning of the next summer term.

- **Preparation:**
  make sure that you have an account for Ilias and the PC-lab for access to course material and practical classes.

Table of Contents

Literature

- Backhaus/Erichson/Plinke/Weiber (2006), *Multivariate Analysemethoden*, 11th edition, Springer Verlag.

- Hair/Black/Babin/Anderson/Tatham (2006), *Multivariate Data Analysis*, 6th edition, Pearson Prentice Hall.

- Sharma (1996), *Applied Multivariate Techniques*, John Wiley & Sons/Inc.

- Delwiche/Slaughter (2006), *The Little SAS Book - A Primer*, SAS Press, 3rd/4th edition.

(1) Introduction

What are Multivariate Methods?

- Multivariate methods can rescue us from **"drowning in information and (being) starved for knowledge"** (Peters, Tom. 1988, Thriving on Chaos. New York: Harper and Row).

- Multivariate methods refer to all **statistical techniques** that simultaneously analyze **multiple variables** and their **relationships** in order to convert data into knowledge.

- Some multivariate methods are already familiar from undergraduate statistics: **bivariate correlation analysis**, **contingency analysis**, **bivariate regression analysis**.

Data and Measurement Scales

The raw material of any multivariate analysis is **data** which results from **measurement processes** that are supposed to attach numbers to given characteristics of objects according to predetermined rules or procedures.

In general, the characteristic considered determines the **measurement accuracy**, i.e. how accurate the characteristic can be expressed in numbers.

EXAMPLE: measurement of the characteristics of a human being

- *body size can easily be measured in numbers*
- *measurement of intelligence/motivation/physical condition quite difficult*

Data and Measurement Scales

Appropriate **measurement scales** that can be used to measure characteristics of objects can be classified into the following four types:

1. nominal scale
2. ordinal scale
3. interval scale
4. ratio scale

presented in increasing order of measurement precision.

Note that the type of measurement scale determines not only (i) the **information content of the data** but also (ii) the **applicability of arithmetic operations.**

Nominal Scale

The **nominal scale** is the most primitive way of measurement: it assigns numbers as a way to **label** or **identify characteristics**. The numbers assigned have no quantitative meaning beyond indicating the presence or absence of the characteristic under investigation.

EXAMPLE: nominal scales

- *gender (1 for male - 2 for female)*
- *religion (1 for catholic - 2 for evangelic - 3 for others)*
- *colour (1 for red - 2 for yellow - 3 for green - 4 for blue)*
- *advertising medium (1 for TV - 2 for newspaper - 3 for poster)*

Valid statistics: statistics based on counts such as mode or frequency distributions. Arithmetic operations are not allowed for.

Ordinal Scale

The **ordinal scale** is the next higher level of measurement precision: it ensures that variables can be **ranked** or **ordered** in relation to the amount of the attribute possessed. The numbers assigned do not indicate the amount of difference between the values.

EXAMPLE: ordinal scales

- *product A is preferred to product B*
- *student A studies harder than student B*
- *school grades*

Valid statistics: mode, median, quantiles and frequency distributions. Arithmetic operations are not allowed for.

Interval Scale

The **interval scale** is the second highest level of measurement precision: it has **constant units of measurement** and an **arbitrary zero point** such that differences between any two adjacent points on any part of the scale are equal, but a point can not be considered to be a multiple of another.

EXAMPLE: interval scales

- *Celsius temperature scale that subdivides the distance between the freezing and boiling point into 100 equally spaced parts, but a temperature of 20°C can not be interpreted as twice as hot as a temperature of 10°C*
- *Fahrenheit temperature scale*

Valid statistics: all except those based on ratios.

Ratio Scale

The **ratio scale** represents the highest form of measurement precision: it possesses the advantages of all lower scales plus an **absolute zero point** (in the sense of non-existent).

EXAMPLE: ratio scales

- *length*
- *weight*
- *height*
- *speed*

There is **no restriction** on the kind of statistics that can be computed for ratio scales data.

Data and Measurement Scales - Remarks

In general, it is possible to **scale down** variables, i.e. to transform a higher to a lower measurement scale, but not the other way around. A scaling down of variables may be useful to increase clarity or to simplify the analysis.

EXAMPLE: scaling down

- *age* $\longrightarrow$ *age groups*
- *income* $\longrightarrow$ *income groups*

Note that such a transformation goes hand in hand with a **loss of information.**

Data and Measurement Scales - Summary

Measurement Scale		Description	Valid arithmetic manipulations
nonmetric	NOMINAL	classification of categorial characteristics	mode, frequency distributions
	ORDINAL	rank with ordinal numbers	mode, frequency distributions, median, quantiles
metric	INTERVAL	constant units of measurement with arbitrary zero point	mode, frequency distributions, median, quantiles, summation, subtraction, mean
	RATIO	constant units of measurement with natural zero point	all arithmetic operations

Classification of Multivariate Methods

We can classify multivariate methods into two main categories:

1. **Dependence methods:**
   identification of one variable or a set of variables as the dependent variable to be predicted or explained by other variables known as independent variables:

		INDEPENDENT VARIABLE	
		metric	nonmetric
DEPENDENT VARIABLE	metric	• regression analysis	• analysis of variance
	nonmetric	• discriminant analysis	• crosstable analysis

   Application requires the existence of a general idea about potential relationships in the data (delivered by theory or on the basis of rational arguments). $\longrightarrow$ **Verification of relationships**.

Classification of Multivariate Methods

2. **Interdependence methods:**
   simultaneous analysis of all variables in the dataset in order to understand or identify why and how the variables are related to each other:

   - principle component analysis
   - factor analysis
   - cluster analysis

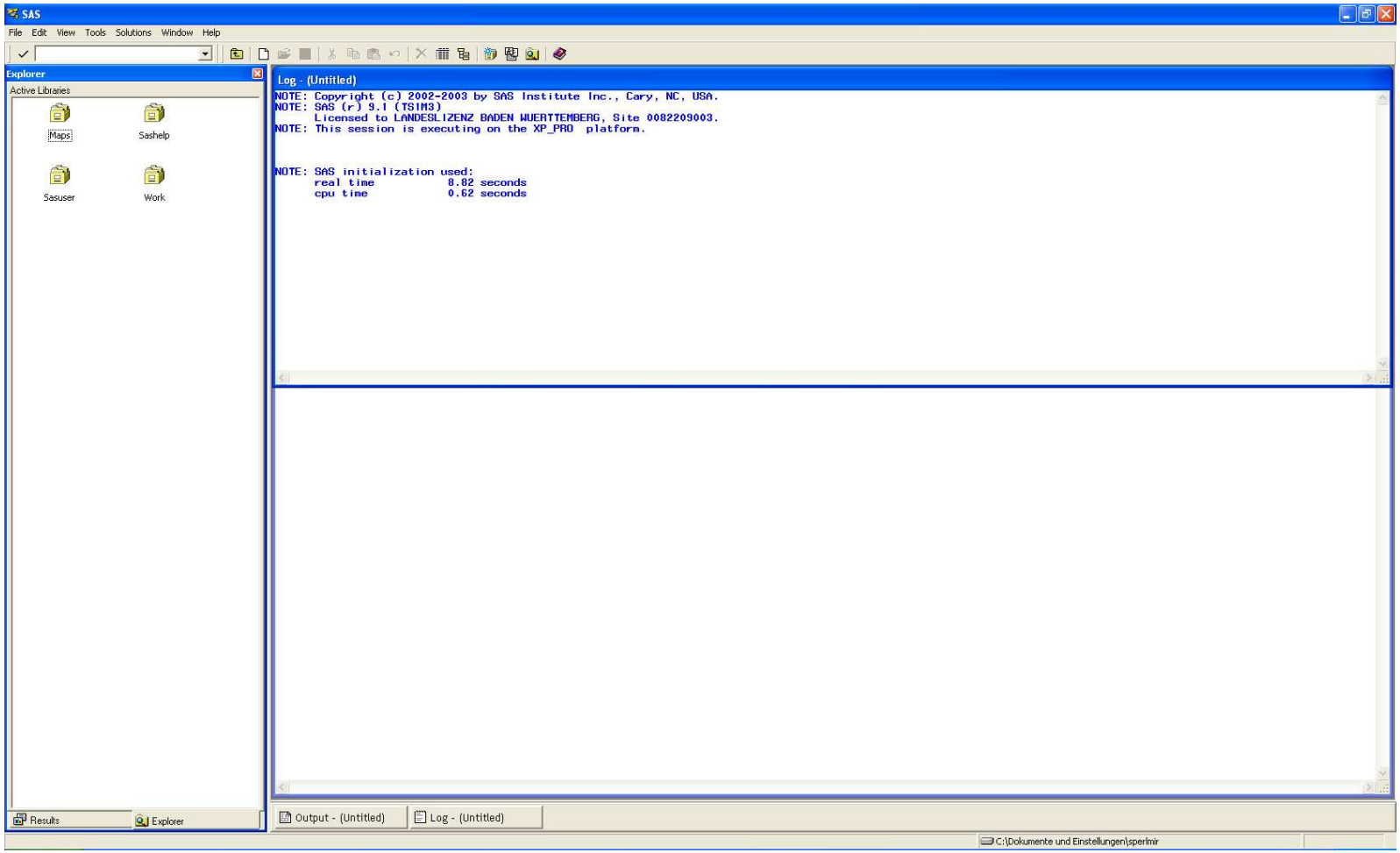
   A-priori general ideas about potential relationships in the data are not required. $\longrightarrow$ **Detection of relationships.**

Introduction to SAS

SAS (pronounced sass)

- both a company, SAS Institute, and software, the SAS System
- started out in the 1970s as software package for statistical analysis
- today SAS offers a variety of products: BaseSAS, SAS/STAT, etc., which are integrated in a single computer program
- Useful online documentation: http://support.sas.com/onlinedoc/913/docMainpage.jsp
- Useful literature: Delwiche/Slaughter (2003), *The Little SAS Book - a primer*, 3rd edition, SAS publishing.

The SAS Microsoft Windows User Interface



The SAS Microsoft Windows User Interface

- **Editor** window for writing/running programs
- **Log** window shows information about the execution of program
  - blue is correct execution (everything is fine)
  - green is a warning (check)
  - red is an error message (find and correct)
- **Output** window shows printed results of any procedure
- **Results** window for navigating around large amounts of procedure output
- **Explorer** window allows to examine contents of SAS data sets and libraries

SAS Datasets

SAS stores data with the help of tables (SAS datasets) comparable to those of Excel. Once your data have been read into a SAS dataset, SAS keeps track of what is where and in what form. Notes on SAS datasets:

- **Variables and observations**: variables in columns, observations in rows
- **Data types:** numeric and character
- **Missings**: missing character data are represented by blanks, missing numeric data by a single period (.)
- **Size of SAS datasets**: up to 32,767 variables, unlimited number of observations as long as your computer's capacity is able to handle and store them

SAS Data Libraries

Before you can use a SAS dataset, you have to tell SAS where to find it. You do that by setting up a SAS library. A SAS library is simply a location (folder/directory/...) where SAS datasets are stored.

The **Active Libraries Window** contains at least three libraries:

- **Sashelp**: information that controls your SAS session along with sample SAS datasets

- **Sasuser**: customizations (such as window size/positioning, colors, fonts, and printer entries) specified for your SAS session

- **Work**: *temporary* storage location/default library where datasets are stored to if you create a dataset without specifying a library. Contents are deleted when closing SAS.

SAS Data Libraries

Datasets of one project are usually stored to a library that you specifically assign for this purpose, e.g. you may organize the data of this course in the library called *mvws08*.

Notes on self-assigned libraries:

- The number of characters for the name of the new library is limited to 8.

- Self-assigned libraries are *permanent*, i.e. contents are not deleted when closing SAS $\neq$ temporary work library ($\rightarrow$ use permanent libraries for datasets that should not be lost, use temporary datasets for dispensable datasets).

Programming Concepts of SAS

SAS is neither menue driven (click - see result) nor command driven (enter command - see result); instead SAS uses statements to write a series of instructions, the so-called SAS program with ingredients:

- **Data steps** to import/modify/reorganize/store the data
- **Procedures** to analyze/estimate/visualize the data
- **SQL** (structured query language) to merge datasets/create tables
- **IML** (interactive matrix language) to conduct matrix operations
- **Macros** to program efficiently

The SAS Program

The SAS program is an ordered set of statements like the set of instructions you use when you go to the bank:

> I would like to make a withdrawal. My account number is 0937. I would like $200. Give me five 20s and two 50s.

You first say what you want to do, then give all the information needed to carry out your request. Here the most important **rules** for writing SAS programs:

- Every SAS statement ends with a semicolon.
- Use names that identify what the data represent, especially for variables.
- Structure your program (e.g. begin a new statement in the next line).
- Comment your code that everyone can read it via /*Comment: ...*/.

Exemplary SAS Program

1_introduction

```
/*************************************************************
******************  INTRODUCTION ********************
*************************************************************
Miriam Sperl
August 2008*/

/*1.3*/
libname mvws08 "C:\Dokumente und Einstellungen\sperlmir\Eigene Dateien\mvws08\sas";

/*2.1*/
data mvws08.stockprices_permanent;
    set mvws08.stockprices;
run;

/*2.2*/
data stockprices;
    set mvws08.stockprices;
run;
```