Discover Patterns in Categorical Time Series using IFS.

Christian H. Weiß

Institute of Applied Mathematics and Statistics

University of Würzburg

This talk is based on the paper

Weiß, C.H., Göb, R.:

Discover Patterns in Categorical Time Series using IFS.

Preprint 260, University of Würzburg, 2005.

Compare conference CD-ROM.

All references mentioned in this talk correspond to the references in this paper.

Iterated Function Systems (IFS)

Background

Iterated function systems (IFS) introduced by Barnsley (1988).

Definition: An *IFS* $\mathcal{F} = \{X;\ c_0, ..., c_{n-1}\}$ consists of

- metric space (X, d) and
- finite set of contractions $c_i : X \to X$, $i = 0, ..., n-1$.

Mapping $c : X \to X$ on metric space (X, d) is *contraction* if for all $x, y \in X$

$$d(c(x), c(y)) \ \leq \ \alpha d(x, y) \qquad \text{with } 0 < \alpha < 1.$$

Chaos algorithm for creating fractals (Barnsley (1988)):

Algorithm: INPUT: *IFS* $\mathcal{F} = \{X;\ c_0, ..., c_{n-1}\}$ and random sequence $i_1, i_2, ...$ with range $\{0, ..., n-1\}$.

1. Initialization: Choose arbitrary $\boldsymbol{x}_0 \in X$.
2. Iteration $t-1 \to t$: Calculate $\boldsymbol{x}_t = c_{i_t}(\boldsymbol{x}_{t-1})$.
3. Repeat step 2.

Iterated Function Systems (IFS)

Applications

Applications of IFS:

- Creating fractals, e.g., Sierpinski triangle, ferns, etc.,
- coding and compression of pictures, Jacquin (1992)
- 'action selection problem for autonomous agents' in robotics, Hoskins (1995)

Applications of IFS (continued):

- synthesis of textures, Chen & Chen (2003)
- analysis and visualization of nucleotide sequences, Jeffrey (1990), Wu et al. (1993)
- sequential pattern analysis. Weiß & Göb (2005)

Analysis of Categorical Time Series

Transformation via Chaos Algorithm

Given: Categorical Time Series $(a_{i_t})_{\mathcal{T}}$ over finite alphabet $\mathcal{A} = \{a_0, ..., a_{n-1}\}$.

Step 1: Transformation of alphabet $\mathcal{A}$ via mapping

$$\boldsymbol{b}: \ \mathcal{A} \to \mathbb{R}^p.$$

Result: Transformed alphabet $\mathcal{A}' = \{\boldsymbol{a}_0, ..., \boldsymbol{a}_{n-1}\} \subseteq \mathbb{R}^p$ with $\boldsymbol{a}_i = \boldsymbol{b}(a_i)$, and time series $(\boldsymbol{Z}_t)_{\mathcal{T}}$ with $\boldsymbol{Z}_t = \boldsymbol{b}(a_{i_t})$.

Step 2: Application of chaos algorithm.

IFS $\mathcal{F} = \{X;\ c_0, ..., c_{n-1}\}$ with linear contractions

$$c_i(\boldsymbol{x}) \;=\; \alpha \boldsymbol{x} \;+\; \beta \boldsymbol{a}_i \qquad \text{with } 0 < \alpha < 1 \text{ and } \beta > 0.$$

Result: Fractal series $\boldsymbol{X}_0 = \boldsymbol{x}_0, \boldsymbol{X}_1, \boldsymbol{X}_2, ...$ with

$$\boldsymbol{X}_t \;=\; \alpha \boldsymbol{X}_{t-1} \;+\; \beta \boldsymbol{Z}_t.$$

Explicitly: For $\boldsymbol{X}_0 := \boldsymbol{0}$

$$\boldsymbol{X}_t \;=\; \beta \cdot \sum_{j=0}^{t-1} \alpha^j \boldsymbol{Z}_{t-j}.$$

Examples:

- $\alpha = \frac{1}{2} = \beta$: *Chaos game representation*, Jeffrey (1990)
- $\alpha = \frac{1}{k} = \beta,\ k \in \mathbb{N}$: *W-curve*. Wu et al. (1993)

Analysis of Categorical Time Series

Main Results

Observation:

If alphabet transformation $\boldsymbol{b}$ and parameter α of IFS are chosen appropriately, then:

- If two segments $(\boldsymbol{Z}_i, ..., \boldsymbol{Z}_{i-k}) = (\boldsymbol{Z}_j, ..., \boldsymbol{Z}_{j-k})$, then points $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ will be *close* in $\mathbb{R}^p$,

- but if the segments differ, then points $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ will be *distant* in $\mathbb{R}^p$.

Idea:

1. Online-transformation of time series $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots$ into fractal series $\boldsymbol{X}_0, \boldsymbol{X}_1, \ldots$.

2. Screen the fractal series $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T, \ldots$ for *close* elements (cluster).

   Points in a cluster correspond to occurrences of similar patterns in time series $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_T$.

Possible realizations:

Transformation of original alphabet $\mathcal{A} = \{a_0, ..., a_{n-1}\}$

- into $\{-1, 1\}^p \subset \mathbb{R}^p$, i.e., edges of p-dim. unit cube
  $\rightarrow$ *cube transformation*,
  see section 5 in Weiß & Göb (2005).

- into unit circle S^1 in $\mathbb{R}^2$, via

$$\boldsymbol{b}(a_k) = (\cos(k \cdot \frac{2\pi}{n}), \sin(k \cdot \frac{2\pi}{n}))^T$$

  $\rightarrow$ *circle transformation*.

Circle Transformations

Background

Let $(\boldsymbol{Z}_t)_{t=1,2,\ldots}$ be series from alphabet $\mathcal{A}' \subset S^1$ resulting from *circle transformation*.

Let $(\boldsymbol{X}_t)_{t=0,1,\ldots}$ be corresponding fractal series

$$\boldsymbol{X}_t \;=\; \alpha \boldsymbol{X}_{t-1} \;+\; \beta \boldsymbol{Z}_t$$

with $0 < \alpha < 1$, $\beta > 0$ and $\boldsymbol{X}_0 = 0$.

For indices $i > j > k > 0$ we compare the segments

$$(\boldsymbol{Z}_i, \ldots, \boldsymbol{Z}_{i-k}) \text{ and } (\boldsymbol{Z}_j, \ldots, \boldsymbol{Z}_{j-k}).$$

Theorem 6.1: Let $d_n = \sqrt{2(1-\cos\frac{2\pi}{n})}$ and $\alpha \leq \frac{d_n}{d_n+3}$.

(iii) If $(\boldsymbol{Z}_i, ..., \boldsymbol{Z}_{i-k}) = (\boldsymbol{Z}_j, ..., \boldsymbol{Z}_{j-k})$, then

$$\|\boldsymbol{X}_i - \boldsymbol{X}_j\| \ \leq \ (2\ (d_n + 3)\ \alpha)\ \frac{\beta}{3}\ \alpha^k.$$

(iv) If $(\boldsymbol{Z}_i, ..., \boldsymbol{Z}_{i-(k-1)}) = (\boldsymbol{Z}_j, ..., \boldsymbol{Z}_{j-(k-1)})$,
but $\boldsymbol{Z}_{i-k} \neq \boldsymbol{Z}_{j-k}$, then

$$\|\boldsymbol{X}_i - \boldsymbol{X}_j\| \ \geq \ d_n\ \frac{\beta}{3}\ \alpha^k.$$

Theorem 6.1 (continued):

Let $d_n \ = \ \sqrt{2(1-\cos\frac{2\pi}{n})}$ and $\alpha \leq \frac{d_n}{d_n+3}$.

(v) If even $\alpha < \frac{d_n}{2(d_n+3)}$, then

$$\|\boldsymbol{X}_i - \boldsymbol{X}_j\| \ < \ d_n \, \frac{\beta}{3} \, \alpha^k,$$

if and only if

$$(\boldsymbol{Z}_i, ..., \boldsymbol{Z}_{i-k}) \ = \ (\boldsymbol{Z}_j, ..., \boldsymbol{Z}_{j-k}).$$

Circle Transformations

An Example

Example: Shakespeare's poem ”Venus and Adonis”.

Ignoring capitals, punctuation marks, and blanks, alphabet $\mathcal{A}$ of $n = 26$ letters:

`a, b, c, d, e, f, g, h, i, j, k, l, m, n,`

`o, p, q, r, s, t, u, v, w, x, y, z.`

Text is a series of letters.

Applying *circle transformation* leads to $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots$.

Example: Shakespeare's poem "Venus and Adonis".

(Continued) Given $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots$.

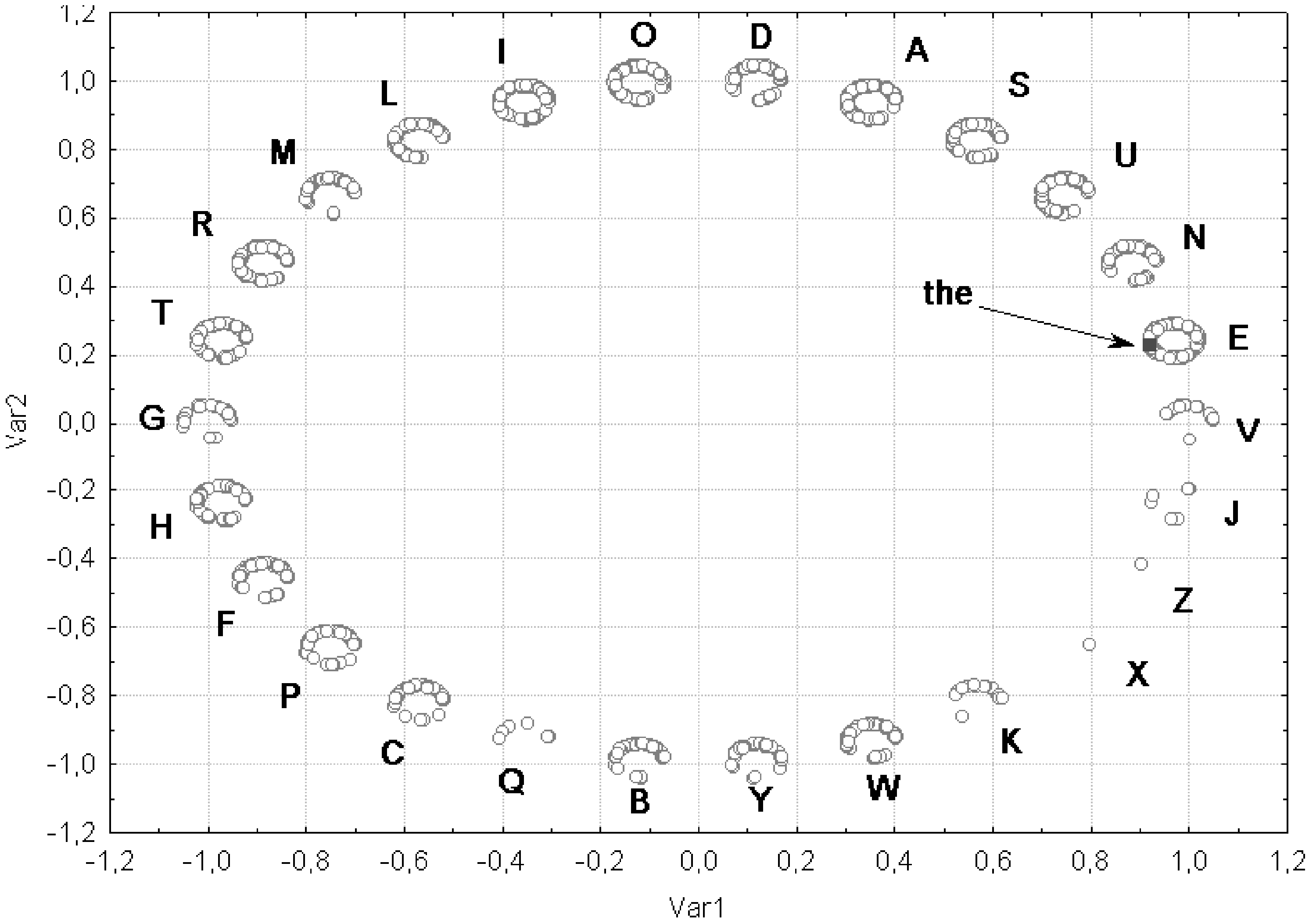
Generation of fractal series $(\boldsymbol{X}_t)_{\mathcal{T}}$ via

$$\boldsymbol{X}_t \;=\; \alpha\boldsymbol{X}_{t-1} \;+\; \beta\boldsymbol{Z}_t, \qquad \text{with } \boldsymbol{X}_0 = \boldsymbol{0},$$

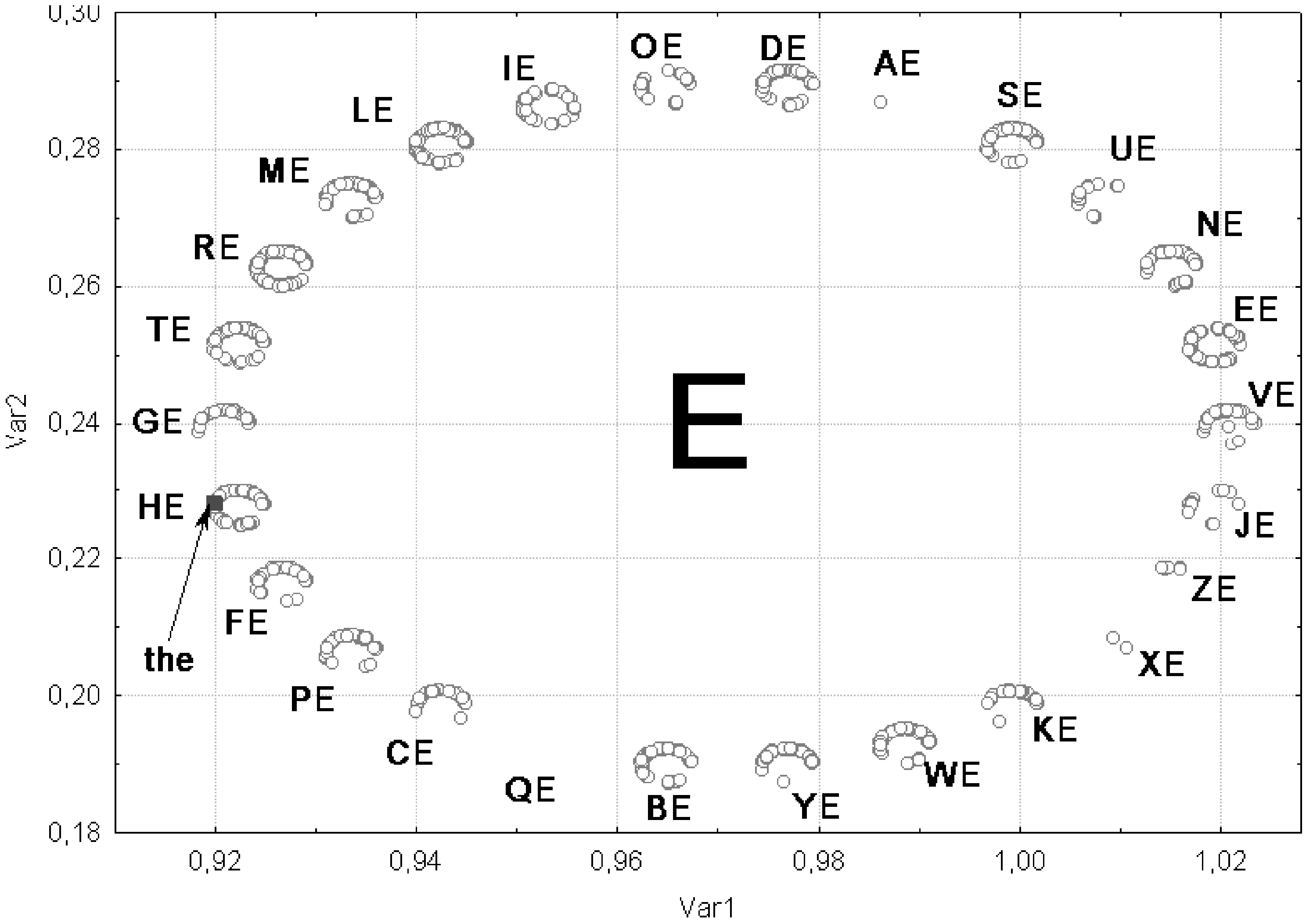
where $\alpha = 0.05$ and $\beta = 1$.

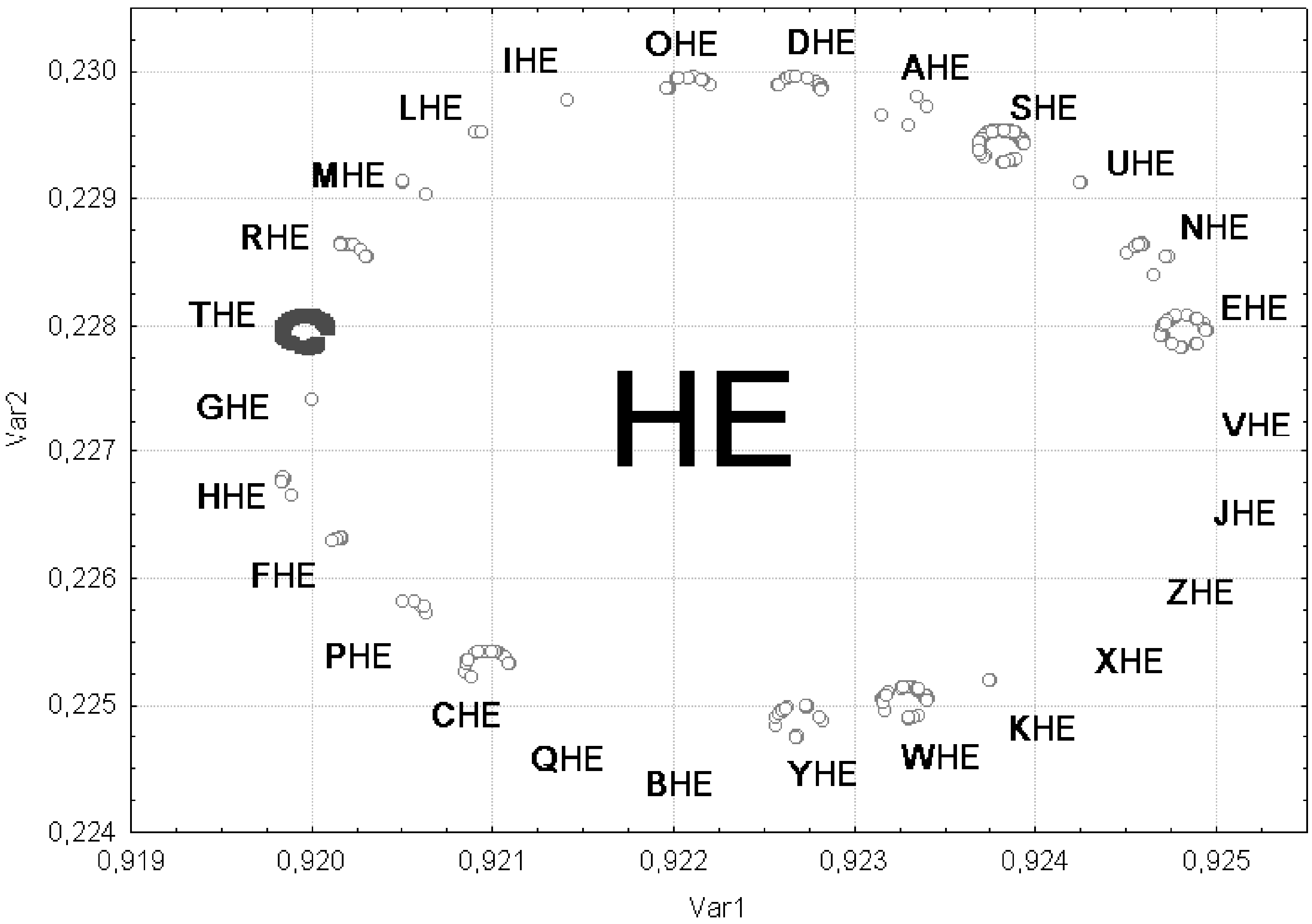
Remark: α even larger than $\frac{d_n}{2(d_n+3)} \approx 0.0372$

$\rightarrow$ pessimistic bound in Theorem 6.1.



1,2
1,0
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,8
-1,0
-1,2
Var2
-1,2
-1,0
-0,8
-0,6
-0,4
-0,2
0,0
0,2
0,4
0,6
0,8
1,0
1,2
Var1
O
D
I
A
L
S
M
U
R
N
the
T
E
G
V
H
J
Z
F
X
P
C
K
Q
B
Y
W


0,30
0,28
0,26
0,24
0,22
0,20
0,18
Var2
0,92
0,94
0,96
0,98
1,00
1,02
Var1
OE
DE
AE
IE
SE
LE
UE
ME
NE
RE
EE
TE
VE
GE
E
HE
JE
the
ZE
FE
XE
PE
KE
CE
WE
QE
BE
YE



HE
IHE
OHE
DHE
AHE
LHE
SHE
MHE
UHE
RHE
NHE
THE
EHE
GHE
VHE
HHE
JHE
FHE
ZHE
PHE
XHE
CHE
KHE
QHE
WHE
BHE
YHE
Var2
0,230
0,229
0,228
0,227
0,226
0,225
0,224
Var1
0,919
0,920
0,921
0,922
0,923
0,924
0,925

Thank You
for Your Interest!

Christian H. Weiß
Institute of Applied Mathematics and Statistics
University of Würzburg