Aalto-yliopisto
Perustieteiden
korkeakoulu

Data envelopment analysis with principal component analysis (final presentation)

Toni Huuhka

27.8.2020

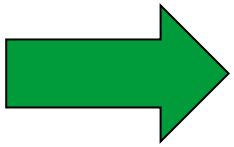
Instructor: *Fabricio Oliveira*

Supervisor: *Fabricio Oliveira*

Systeemianalyysin
laboratorio

Background

- The majority of companies are not operating in an ideal manner
- How could a company improve its operational efficiency?



1. Rank similar companies and identify the most efficient ones
2. Identify the factors that cause the difference in efficiency
3. Implement these ”best-practices”

Contents

- Data
- Methods
  - Principal Component Analysis
  - Data Envelopment Analysis
- Results
- Conclusions

Data

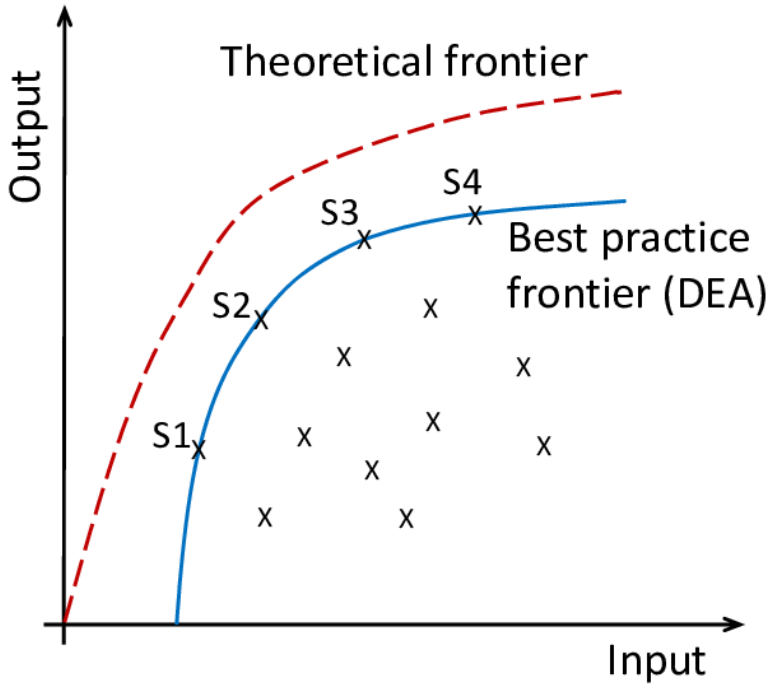
- 221 active Scandinavian retailers
  - No retailers of motor vehicles and motorcycles
  - Revenue ≥ 50 million €
- Data retrieved from Orbis Database
- 3 input variables, 2 output variables
  - Inputs: *Number of Employees* (I1), *Number of current Directors & Managers* (I2), *Number of current Advisors* (I3)
  - Outputs: *Sales* (O1), *Cost of Employees* (O2)

Data: Correlations

	O1	O2	I1	I2	I3
O1	1.0000				
O2	0.9573	1.0000			
I1	0.9173	0.9676	1.0000		
I2	0.4708	0.5245	0.4931	1.0000	
I3	0.1576	0.1733	0.1845	0.2295	1.0000

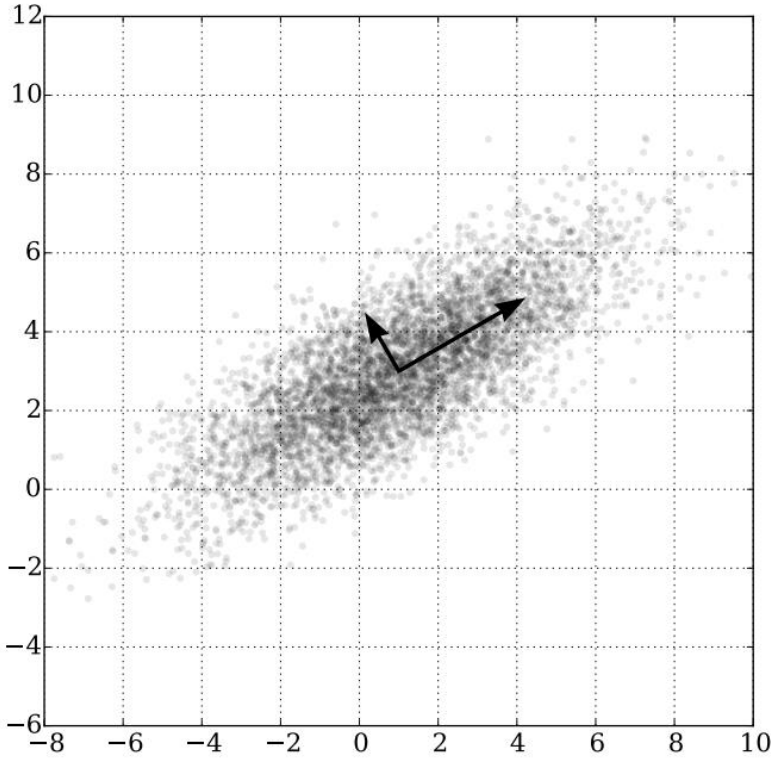
Methods: Data Envelopment Analysis

- Linear programming method
- Identifies a ”best-practice frontier”
- Sensitive to outliers



Methods: Principal Component Analysis

- A method for creating uncorrelated basis vectors
- Commonly used for variable reduction
- A more robust version
  - *Robust PCA*



Methods: Principal Component Analysis

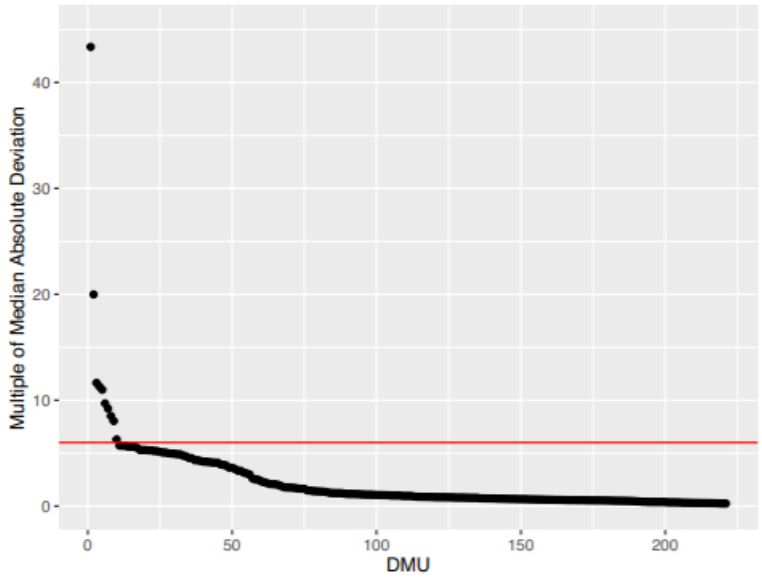
- Both DEA and PCA are sensitive to outliers
- The effects of centering and scaling the data before PCA were studied
- PCA-DEA and RPCA-DEA was performed with every combination of scaling and centering

A total of six efficiency scores were computed in addition to the original three

Results: PCA and Robust PCA

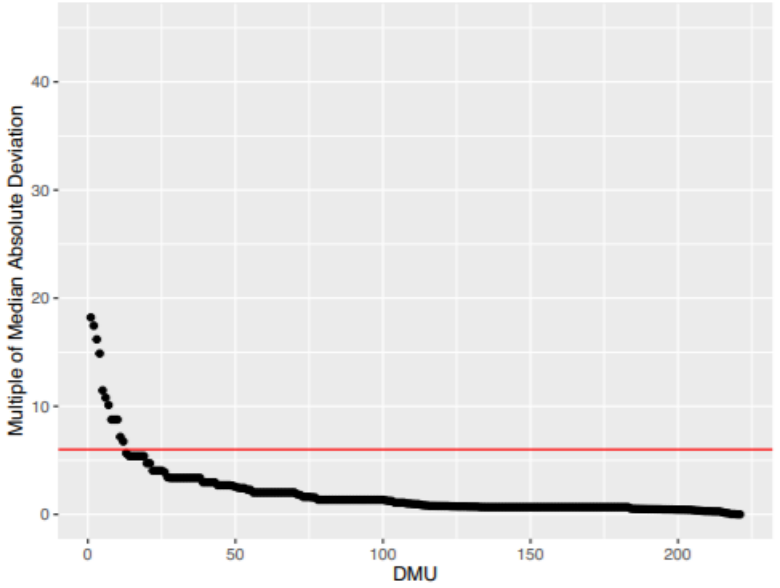
PCA

	IPC1	IPC2	IPC3	OPC1	OPC2
R^2	0.543	0.289	0.168	0.979	0.021
Cum. R^2	0.543	0.832	1.000	0.979	1.000



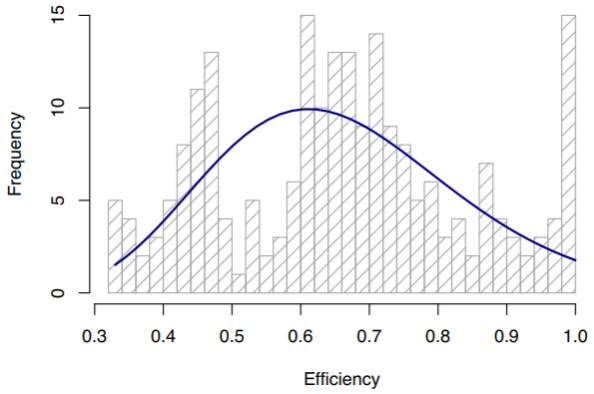
Robust PCA

	IPC1	IPC2	IPC3	OPC1	OPC2
R^2	0.534	-0.087	0.129	0.978	-0.460
Cum. R^2	0.534	0.447	0.576	0.978	0.519

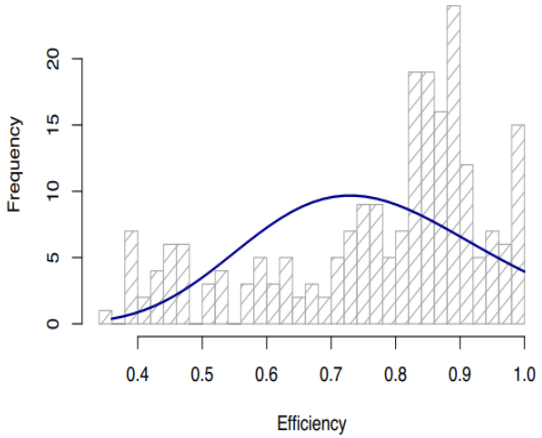


Results: DEA, PCA-DEA, RPCA-DEA

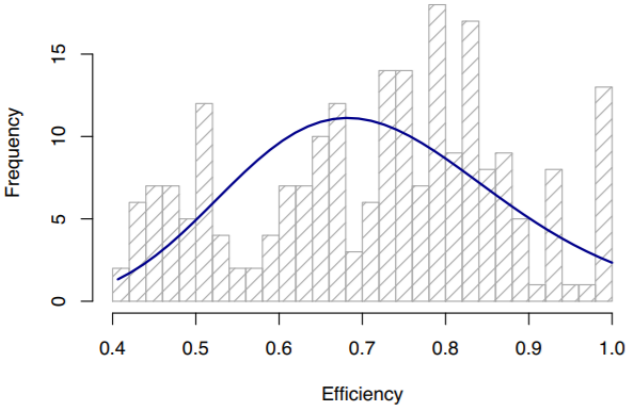
DEA



PCA-DEA



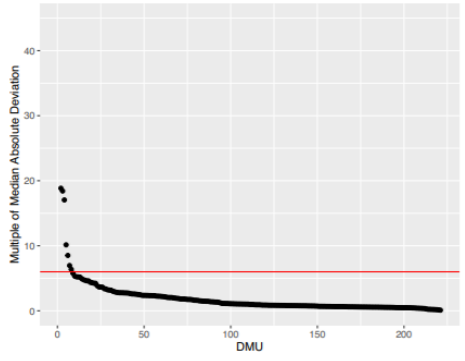
RPCA-DEA

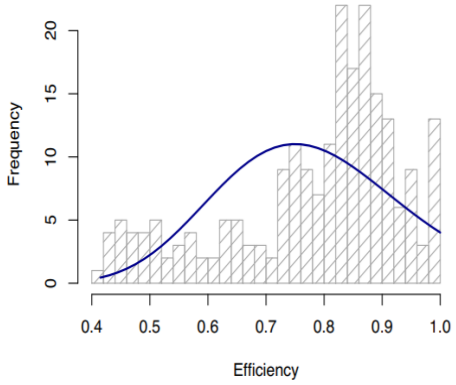


Results: Non-centered PCA-DEA & RPCA-DEA

PCA-DEA

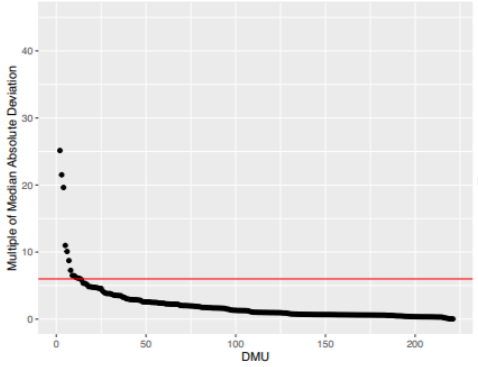
	IPC1	IPC2	IPC3	OPC1	OPC2
R^2	0.735	0.192	0.073	0.982	0.018
Cum. R^2	0.735	0.927	1.000	0.982	1.000

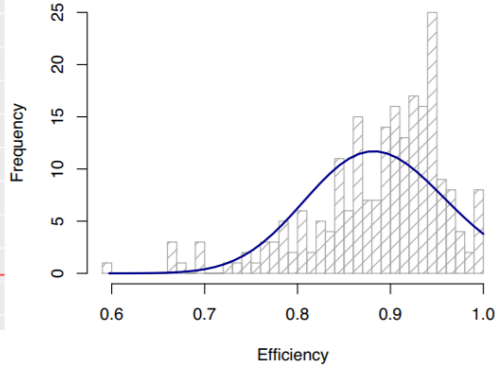




RPCA-DEA

	IPC1	IPC2	IPC3	OPC1	OPC2
R^2	0.785	0.122	-0.397	0.978	-0.404
Cum. R^2	0.785	0.907	0.510	0.978	0.574

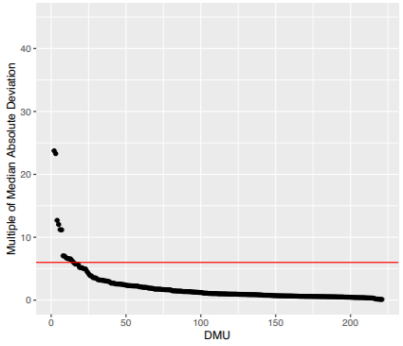


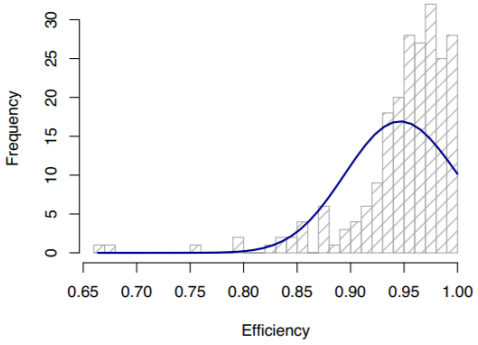


Results: Non-scaled PCA-DEA & RPCA-DEA

PCA-DEA

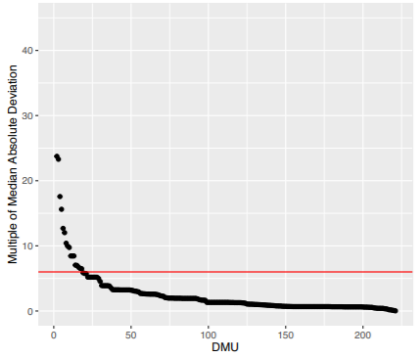
	IPC1	IPC2	IPC3	OPC1	OPC2
R^2	1.000	0.000	0.000	0.999	0.001
Cum. R^2	1.000	1.000	1.000	0.999	1.000

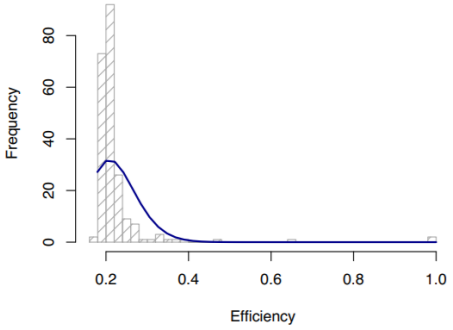




RPCA-DEA

	IPC1	IPC2	IPC3	OPC1	OPC2
R^2	1.000	0.000	0.000	0.999	-0.013
Cum. R^2	1.000	1.000	1.000	0.999	0.986

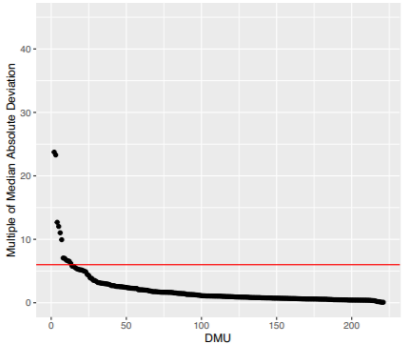


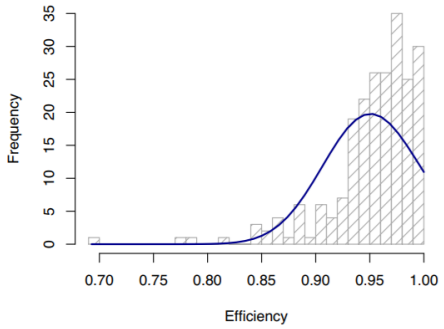


Results: Non-scaled, non-centered PCA-DEA & RPCA-DEA

PCA-DEA

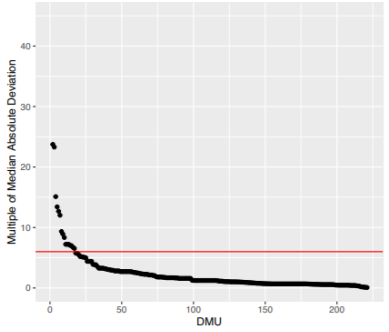
	IPC1	IPC2	IPC3	OPC1	OPC2
R^2	1.000	0.000	0.000	0.999	0.001
Cum. R^2	1.000	1.000	1.000	0.999	1.000

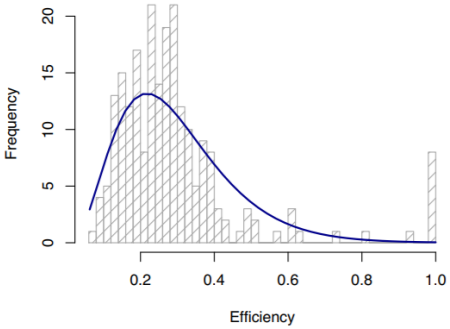




RPCA-DEA

	IPC1	IPC2	IPC3	OPC1	OPC2
R^2	1.000	0.000	0.000	0.999	-0.014
Cum. R^2	1.000	1.000	1.000	0.999	0.985





Conclusions

- Scaling more important than centering prior to any form of (R)PCA-DEA
- Robust estimators improve the results of PCA-DEA
- No clear advantage was found between centered and not-centered data with either PCA-DEA or RPCA-DEA

Future prospects

- The effects of removing outliers should be studied, the outliers were not removed in this study
- Different assumptions regarding returs to scale could be investigated
- Additional research regarding when variable reduction techniques are necessary to implement