

Smarte Datenintegration durch Benchmarking-as-a-Service – Vorschlag eines Vorgehensmodells zur Entwicklung eines unternehmensübergreifenden Informationssystems mittels Data Envelopment Analysis

Andreas Varwig¹, Friedemann Kammler² und Oliver Thomas³

Abstract: Obwohl Maschinen- und Anlagenbauer immer komplexere Sensortechnologien und Informationssysteme verbauen, haben sie nur selten Zugriff auf detaillierte Laufzeitdaten. Anlagenbetreiber, die oftmals die Eigentümer der Daten sind, verweigern ihnen den Zugriff, während sie selbst nicht über die Mittel verfügen, die Daten auszuwerten. In der Folge entstehen bei jedem Betreiber unerschlossene Data Lakes. In diesem Artikel wird ein Vorgehensmodell zur Erschließung und betreiberübergreifenden Integration der ökonomisch verwertbaren Informationen, basierend auf der Data Envelopment Analysis, vorgestellt.

Keywords: Smart Data Linking; Data Envelopment Analysis; Industrie 4.0; Information Systems

1 Einführung

Die zunehmende Digitalisierung führt zu fundamentalen Veränderungen in der industriellen Wertschöpfung. Dabei steht der Maschinen- und Anlagenbau vor besonderen Herausforderungen. Obwohl Anlagenbauer immer umfangreichere Sensortechnologien und IT-Systeme zur Protokollierung von Maschinenbetriebsdaten verbauen, haben sie nur eingeschränkten Zugriff auf die Informationen. Theoretisch könnten Sie diese neuen detaillierten Laufzeitinformationen u.a. dazu verwenden, die Maschinenkonstruktion zu optimieren und neue Dienstleistungen anzubieten. Die Daten gehören jedoch zumeist dem Betreiber einer Anlage, welcher den Zugang verweigert. In der Folge entsteht eine Vielzahl brachliegender Data Lakes (DL) bei den Anlagenbetreibern.

Wir zeigen, wie Maschinenhersteller durch den Einsatz von quantitativen Datenanalysemethoden ihr Dienstleistungsspektrum erweitern, ökonomisch wertvolle Informationen aus dezentralen DL identifizieren und Anlagenbetreibern gleichzeitig einen Anreiz zum Teilen ihrer Maschinendaten liefern können. Hierfür skizzieren wir ein Vorgehensmodell, welches Methoden der Data Envelopment Analysis (DEA) nutzt, um Empfehlungen für

¹ Universität Osnabrück, Fachgebiet Informationsmanagement und Wirtschaftsinformatik, Katharinenstraße 3, 49074 Osnabrück, andreas.varwig@uos.de

² Universität Osnabrück, Fachgebiet Informationsmanagement und Wirtschaftsinformatik, Katharinenstraße 3, 49074 Osnabrück, friedemann.kammler@uos.de

³ Universität Osnabrück, Fachgebiet Informationsmanagement und Wirtschaftsinformatik, Katharinenstraße 3, 49074 Osnabrück, oliver.thomas@uos.de

die optimale Konfiguration von Maschinen und Anlagen abzuleiten. Der Artikel ist folgendermaßen aufgebaut: Zunächst wird der konkrete Handlungsbedarf am Beispiel eines Herstellers von industriellen Druckanlagen beschrieben. Im dritten Abschnitt werden das methodische Vorgehen der DEA und ein kurzer Literaturüberblick dargestellt. Der vierte Abschnitt skizziert das Vorgehensmodell zur Entwicklung eines betreiberübergreifenden Informationssystems. Die resultierenden Implikationen und Limitationen werden im fünften Abschnitt diskutiert. Der Artikel schließt mit einer Zusammenfassung und einem Ausblick auf nachfolgende Forschungsarbeiten.

2 Ungenutzte Data Lakes im Maschinen- und Anlagenbau

Im Maschinen- und Anlagenbau herrscht Informationsasymmetrie. Der eingeschränkte Zugriff auf Betriebsdaten verkaufter Maschinen behindert die Hersteller darin, ihre Produkte und Dienstleistungen zu verbessern und neue Geschäftsmodelle zu erschließen. Die Problematik lässt sich am Beispiel eines Herstellers für industrielle Druckanlagen verdeutlichen. Die Anlagen dieses Unternehmens bestehen aus tausenden Einzelteilen und Modulen. Zur Überwachung und Analyse des Maschinenzustands werden seit Jahren immer umfangreichere Sensoren verbaut. Diese Sensoren ermöglichen es u.a. Temperaturen, Vibrationen, akustische Auffälligkeiten und Luftzusammensetzungen an unterschiedlichen Positionen der Anlage zu erfassen. Gleichzeitig werden auch die IT-Komponenten und Steuerungshilfen der Anlage immer komplexer und detaillierter. Sie erfassen während Laufzeit u.a. Informationen über die verarbeiteten Materialien, aufgetretene Fehlermeldungen, Materialdurchlaufgeschwindigkeiten, und -spannungen und Anschlagwinkel von Federungen und Walzen in einzelnen Modulen der Druckwerke. Das Ziel des Anlagenbauers ist es, die Informationen zu nutzen, um das Maschinendesign zu verbessern, Wartungs- und Ersatzteilbedarfe frühzeitig zu erkennen oder komplett neue datenbasierte Dienstleistungen anzubieten.

Die erzeugten Betriebsdaten sind jedoch das Eigentum der Anlagenbetreiber. In den meisten Fällen sind diese Unternehmen nicht in der Lage, die entstandenen Informationen nutzenbringend weiterzuverarbeiten. Teils fehlen die notwendigen IT-Infrastrukturen oder es ist kein Methodenwissen zur statistischen Datenverarbeitung im Unternehmen verfügbar. Auch das schlichte Fehlen von Use Cases wurde in Kundenbefragungen als Ursache ausgemacht. Dennoch verweigern die Betreiber dem Anlagenhersteller den Datenzugriff und archivieren die Betriebsinformationen in lokalen Speichersysteme.⁴

⁴ Aufgrund ihrer Beschaffenheit im Hinblick auf ihr Volumen, die Vielseitigkeit, die teils sehr hohe Frequenz und den (unerschlossenen) Wert der Daten, lässt sich die vorliegende Situation als Big Data Herausforderung beschreiben, siehe hierzu u.a. [OFT14].

3 Funktionsweise und Anwendung der Data Envelopment Analysis

Die Data Envelopment Analysis wurde 1978 von [CCR78] eingeführt und seitdem stetig weiterentwickelt. Heute beschreibt der Begriff ein breites Portfolio an quantitativen Methoden zum Benchmarking in komplexen Leistungserstellungsprozessen. Die Kernidee der DEA ist es, einen objektiven, methodischen Effizienzvergleich in einer heterogenen Gruppe von Leistungserstellern, sog. DMU (von engl. Decision Making Units), durchzuführen. Dabei können alle leistungsrelevanten Informationen (i.S.v. Inputs und Outputs), die numerisch abbildbar sind, berücksichtigt werden. Während der Analyse werden diese Werte für jede zu vergleichende DMU zu einer einzigen Leistungskennzahl, dem sog. Effizienzparameter, verdichtet.

3.1 Methodisches Vorgehen

Das Lineare Optimierungsproblem (LP) des Basismodells von [CCR78] ist in den Gleichungen (1) - (4) dargestellt. Es ist für jede DMU der zu vergleichenden Gruppe zu lösen.

$$\text{Min}_{\lambda} \theta_i, \text{ s. t.} \quad (1)$$

$$X\lambda_i \leq \theta_i x_i \quad (2)$$

$$Y\lambda_i \geq y_i \quad (3)$$

$$\lambda_i \geq 0, \theta_i \geq 0 \quad (4)$$

Bei der Leistungsbewertung wird ermittelt, welchen Ressourcenaufwand (X) eine DMU mindestens einsetzen müsste, um ihren aktuellen Output (Y) zu erzielen. Dies geschieht durch die Bildung von Linearkombinationen aller zu vergleichenden DMUs. Die Relation aus optimalem Ressourceneinsatz zu aktuellem Ressourceneinsatz wird als Effizienzparameter θ_i bezeichnet. Entspricht θ_i beispielsweise 0.80, so könnte die betrachtete DMU i den aktuellen Output bereits mit 80% der aktuell eingesetzten Ressourcen erzielen.

Gleichung (1) ist die Zielfunktion des LP. Sie lässt sich mit der Aufforderung: „Finde das minimale Inputniveau, welches zur Erstellung des aktuellen Outputs notwendig ist!“ übersetzen. Die Gleichungen (2) - (4) definieren die Nebenbedingungen, welche dabei zu beachten sind: (2): Das optimale Inputniveau muss kleiner oder gleich dem aktuellen Ressourceneinsatz sein, (3): Unter Verwendung des optimalen Ressourceneinsatzes wird mindestens derselbe Output erzeugt wie vor der etwaigen Verhaltensanpassung, (4): Nur positive Linearkombinationen anderer existierender Input-Output-Profile sind zur Identifikation der individuell-optimalen Lösung zulässig.⁵

⁵ Für eine Einführung in die Grundlagen verschiedener DEA-Methoden siehe bspw. [CST07] und [CZS04].

Der λ_i -Vektor ist das zweite zentrale Ergebnis einer DEA. Er zeigt auf, zu welchem Anteil eine DMU für die Konstruktion eines jeweiligen individuellen Referenzprofils verwendet wurde. Alle DMUs, welche anteilig zumindest bei der Konstruktion eines Referenzprofils für eine andere DMU verwendet wurden, stellen eine Best-Practice dar. Sie sind Blaupausen, anhand derer sich ineffiziente DMUs zur Leistungsverbesserung orientieren können.

3.2 Stand der Forschung und DEA-Applikationen

Seit ihrer Einführung wurde die DEA für die Untersuchung unterschiedlichster Fragestellungen eingesetzt. [MSS03] verwenden sie, um die Leistung deutscher Arbeitsämter zu bewerten. Eine Auswertung der Effizienz von Gerichtsverhandlungen ist bei [Sc04] zu finden. Auch in der Privatwirtschaft ist eine Vielzahl an Anwendungen entstanden. Insbesondere für die Analysen von Finanzdienstleistern und -dienstleistungen erfreut sich die DEA großer Popularität.⁶ Aktuelle Untersuchungen bieten u.a. [VVP13] oder [VVL17]. [SPS05] setzen Methoden der DEA zur Priorisierung und Modifikation von Projekten zur Implementierung von Informationssystemen im Bankwesen ein. [Th96] untersuchen die Optimierungsmöglichkeiten in der Erdölförderung. Daneben ist die Analyse der Erzeugung und Bereitstellung von Elektrizität ein wiederkehrender Forschungsgegenstand in der DEA-Literatur, siehe u.a. [ALS99]. Neuere Methodenentwicklungen und Applikationen beschäftigen sich speziell mit der Erschließung und Verarbeitung von Big Data.⁷ So untersuchen u.a. [Sol6] die Eignung der DEA in der Verarbeitung von Big Data zur Einschätzung regionaler Umweltbelastung.

4 Benchmarking-as-a-Service zur Datenintegration

Wir haben ein Vorgehensmodell, bestehend aus 6 Schritten und basierend auf der DEA, entwickelt, mit dem es möglich ist, eine betreiberübergreifende Informationsplattform zu entwickeln. Durch den Einsatz der DEA sind Betreiber in der Lage, ein umfassendes Benchmarking ihrer Produktion vorzunehmen. So können sie Maschinenkonfigurationen identifizieren, welche Ineffizienzen verursachen. Damit werden Informationen ermittelt, aus denen auch andere Anlagenbetreiber lernen könnten. Durch die Integration dieser Daten kann eine Datenbasis geschaffen werden, aus der alle Betreiber Konfigurationsempfehlungen für einzelne Maschinen erhalten können. Die Datenintegration durch DEA-basiertes Benchmarking-as-a-Service (BaaS) kann somit ein neues Geschäftsmodell für den Anlagenbauer darstellen, welches den Zugriff auf bislang unerreichbare Maschinenlaufzeitinformationen ermöglicht. Die 6 Schritte des Vorgehensmodells sind in Abb. 1 dargestellt.

⁶ Allein zur Bankeneffizienz wurden bislang weit mehr als 200 Studien durchgeführt und publiziert.

⁷ In einer Analyse von mehr als 2000 Publikationen zu Methodenerweiterungen und Anwendungen der DEA, die in den Jahren 2010-2014 entstanden sind, zeigen [LLL16] die aktuellen Herausforderungen der DEA-Forschung auf.

4.1 Vorgehensmodell

Im ersten Schritt des Vorgehensmodells sind Maschinendaten zu identifizieren und zu extrahieren, anhand derer sich die erbrachte Leistung einer Maschine aus Sicht eines Betreibers bewerten lässt. Als Inputs kommen alle Werte in Frage, die während der Leistungserstellung möglichst geringgehalten werden sollen. Hier bieten sich bspw. die Produktionsdauer, die verwendeten Produktionsmaterialien oder die Anzahl an Fehlermeldungen während eines Produktionslaufes an. Outputs sind solche Indikatoren, die während der Produktion zu steigern sind. Beispiele hierfür sind die produzierte Menge und die durchschnittliche Qualität. Im nächsten Schritt (2) werden die Leistungen der einzelnen Maschinendurchläufe unter der Berücksichtigung der individuellen Selektion relevanter Inputs und Outputs durch eine DEA analysiert. Die Analyse kann sowohl direkt auf der IT-Architektur der Maschine, bspw. unter Verwendung eines R-Skriptes, als auch auf lokalen Systemen des Betreibers durchgeführt werden. Hierbei werden die Durchläufe in effiziente und ineffiziente Durchläufe eingeteilt. Alle vorhandenen Informationen zu den effizienten Durchläufen werden nun (Schritt 3) an den Maschinenhersteller übertragen. Dies beinhaltet auch Informationen, die nicht bei der Leistungsbewertung betrachtet wurden. Der Maschinenhersteller erhält so auch die Maschinenlogs, Konfigurationsinformationen und hochfrequente Sensordaten.

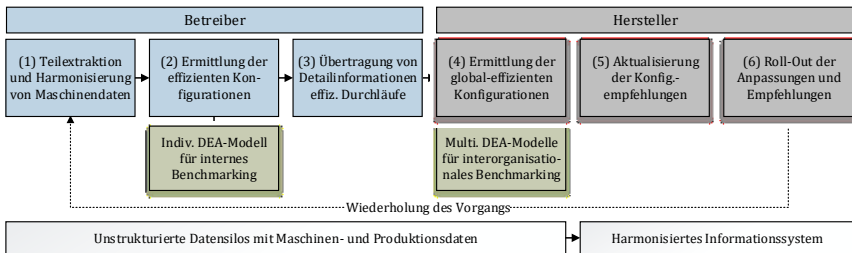


Abb. 1: Vorgehensmodell zur Datenintegration durch Benchmarking-as-a-Service

Im vierten Schritt analysiert und vergleicht der Anlagenbauer die Produktionsprozesse aller Betreiber. Hierbei kann es notwendig sein, die jeweils individuellen Leistungsprozesse der einzelnen Hersteller für den Gesamtbestand der Daten zu modellieren. Dementsprechend wird in Schritt 4 nicht zwingend nur ein Leistungsvergleich durchgeführt. Die effizienten Leistungen der jeweiligen Betreiber werden nun gemäß dem individuellen Leistungsverständnis miteinander verglichen. So kann identifiziert werden, ob ein Maschinendurchlauf auch im betreiberübergreifenden Vergleich nach dem eigenen Leistungsverständnis noch als effizient eingestuft wird, oder ob sich Verbesserungspotentiale ableiten lassen. Während dieser Analyse können auch historische Leistungsdaten aus den bisherigen Konfigurationsempfehlungen des Herstellers berücksichtigt werden. Im fünften Schritt aktualisiert der Maschinenhersteller seine Konfigurationsempfehlungen. Dabei wird über die Zeit ein Informationssystem aufgebaut, welches die individuellen Leistungsbegriffe, unterschiedliche Lastprofile und effiziente Konfigurationen historisiert. Im letz-

ten Schritt erhalten die Betreiber eine Rückmeldung über die betreiberübergreifende Effizienz ihrer Produktion und ggf. Empfehlungen darüber, wie die Leistung durch Anpassungen der Maschinenkonfiguration gesteigert werden kann. Auch werden die aktualisierten Konfigurationsempfehlungen für die unterschiedlichen Leistungsprozesse und Lastprofile an die Betreiber ausgeliefert.

4.2 Numerisches Beispiel

Die ersten drei Schritte des Vorgehensmodells können anhand eines numerischen Beispiels verdeutlicht werden. Hierzu werden zufällig Daten zu 12 Durchläufen einer Maschine erzeugt. Die Inputs sind gegeben durch die Produktionsdauer (D), die Materialkosten (K) und die Häufigkeit von kritischen Fehlermeldungen (F). Als Outputs werden die Produktionsmenge (M) und die durchschnittliche Produktionsqualität (Q) modelliert. Die Analyse der Beispieldaten erfolgt unter Verwendung der freien Skriptsprache R.

```
#DEA-Skript
library(Benchmarking);
Data <- read.csv2(paste0(getwd(), "/data.csv"));
E <- dea(Data[,1:3],Data[,4:5],RTS="vrs", SLACK = "TRUE")
out <- data.frame(Data,E$lambda,E$eff,E$slack,E$ss,E$sy)
```

In Abweichung zum Basismodell wird der Effizienzwert eines jeweiligen Druckdurchlaufs im Beispiel unter der Annahme variabler Skalenerträge berechnet. Dies ermöglicht es Laufzeiteffekte, wie bspw. das Wegfallen von Warmlaufzeiten, zu berücksichtigen. Die Ausgangsdaten und die Ergebnisse sind nachfolgend in Tabelle 1 dargestellt.

Lf.	D	K	F	M	Q	θ	λ_1	λ_3	λ_5	λ_6	λ_7
1	20	36	11	52	28	1	1	0	0	0	0
2	48	78	8	10	4	.53	.15	0	0	.85	0
3	46	75	5	37	40	1	0	1	0	0	0
4	24	42	6	4	20	.96	.07	0	.45	.49	0
5	22	39	8	45	32	1	0	0	1	0	0
6	24	42	3	12	8	1	0	0	0	1	0
7	88	138	8	60	16	1	0	0	0	0	1
8	24	42	11	10	16	.88	.84	0	0	.16	0
9	66	105	8	4	40	.71	0	1	0	0	0
10	44	72	6	12	24	.75	0	.39	.15	.46	0
11	70	111	11	3	16	.39	0	.07	.24	.69	0
12	42	69	8	1	24	.68	0	.19	.41	.40	0

Tab. 1: Numerisches Beispiel einer DEA-Anwendung in Schritt 2 des Vorgehensmodells

Die Analyse identifiziert 5 der 12 Maschinendurchläufe als effizient (grau hinterlegt). Diese sind die Durchläufe 1,3,5,6 und 7. Ihr berechneter θ -Wert entspricht 1. Dies deutet darauf hin, dass der erzeugte Output nicht mit weniger als den eingesetzten Ressourcen zu

erzielen gewesen wäre. In den λ -Spalten von Tabelle 1 ist der Anteil angegeben, zu welchem ein effizienter Durchlauf für die Bestimmung der Referenzprofile der ineffizienten Durchläufe verwendet wird. Entsprechend Schritt 4 des skizzierten Vorgehensmodells werden nun alle Datensätze, die während dieser Durchläufe entstanden sind, auf an das zentrale Informationssystem übertragen.

5 Implikationen und Limitationen

Die Methoden der DEA sind sehr flexibel. Sie können zumeist unabhängig von den Maßeinheiten der verfügbaren Leistungsinformationen verwendet werden. Gleichzeitig können somit etwa monetäre Größen (bspw. Rohstoffausgaben) und metrische Informationen (bspw. die verfügbare Verkaufsfläche) oder simultan kardinal- und ordinalskalierte Informationen (qualitative Leistungsindikatoren) berücksichtigt werden, siehe u.a. [CZ06] oder [Sa07]. Damit ist die DEA geeignet, um die ökonomisch verwertbaren Informationen aus den bislang nicht erschlossenen, lokalen Data Lakes der Anlagenbetreiber zu gewinnen.

Obwohl das Vorgehen generalisierbar und auf weitere Branchen übertragbar ist, ist die Modellierung des Leistungsprozesses eine individuell zu lösende Herausforderung. Es ist nicht davon auszugehen, dass die zu untersuchenden Leistungsinformationen in anderen Szenarien bereits vollständig und strukturiert vorliegen. Selbst im beschriebenen Beispiel des Druckanlagenherstellers wurde das Vorgehensmodell bislang nicht in der Praxis getestet. Es bleibt abzuwarten, ob die Aussicht auf eine verbesserte Maschinenkonfiguration, verringerte Stand- und Ausfallzeiten und eine verbesserte Fertigungsqualität die Betreiber von der Bereitstellung ihrer Laufzeitinformationen überzeugen kann.

6 Zusammenfassung und Ausblick

Obwohl detaillierte Leistungs- und Betriebsinformationen zu ihren Anlagen eigentlich verfügbar sind, haben viele Maschinenbauer keinen Zugang zu diesen Daten. Betriebsdaten sind in der Regel das Eigentum des Anlagenbetreibers. Am Beispiel der Situation eines Herstellers von industriellen Druckanlagen wurde ein Vorgehensmodell vorgestellt, welches es ermöglicht, historische Produktionsdaten in für den weiteren Erkenntnisgewinn relevante und irrelevante Datensätze einzuteilen. Durch die Freigabe der so identifizierten relevanten Informationen kann der Hersteller Empfehlungen zur Optimierung der Maschinenkonfiguration und produktionszielabhängige Standardkonfigurationen bereitstellen.

In nachfolgenden Forschungsarbeiten gilt es, das Vorgehensmodell in der Praxis umzusetzen und zu evaluieren. Hierfür sind u.a. konkrete und generalisierbare Produktionsprozesse, im Sinne von zu betrachtenden Inputs und Outputs, zu definieren. Auch sind die Anforderungen an die technische Umsetzung des Informationssystems zu erheben. Nicht zuletzt ist für die Bewertung des Gesamtnutzens des Konzepts eine erwartbare Kosten- und Ertragskalkulation zu erstellen.

Literaturverzeichnis

- [ALS99] Athanassopoulos, A., Lambroukos, N., & Seiford, L. (1999). Data envelopment scenario analysis for setting targets to electricity generating plants. *European Journal of Operational Research*, 115, 413–428.
- [CCR78] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.
- [CZ06] Cook, W. D., & Zhu, J. (2006). Rank order data in DEA: A general framework. *European Journal of Operational Research*, 174(2), 1021–1038.
- [CST07] Cooper, W., Seiford, L., & Tone, K. (2007). *Data Envelopment Analysis - A Comprehensive Text with Models, Applications, References and DEA-Solver Software* (2nd edn.). Springer.
- [CZS04] Cooper, W., Zhu, J., & Seiford, L. M. (2004). *Handbook on Data Envelopment Analysis* (2nd ed.). Kluwer Academic Publishers.
- [LLL16] Liu, J. S., Lu, L. Y. Y., & Lu, W.-M. (2016). Research fronts in data envelopment analysis. *Omega*, 58, 33–45.
- [MSS03] Mosley, H., Schütz, H., & Schmid, G. (2003). *Effizienz der Arbeitsämter: Leistungsvergleich und Reformpraxis*. Berlin: Ed. Sigma.
- [OFT14] Özcan, D., Fellmann, M., & Thomas, O. (2014). Towards a Big Data-based Technical Customer Service Management. 44. Jahrestagung Der Gesellschaft Für Informatik – INFORMATIK 2014, 187–198.
- [Sa07] Saen, R. F. (2007). Suppliers selection in the presence of both cardinal and ordinal data. *European Journal of Operational Research*, 183(2), 741–747.
- [Sc04] Schneider, M. (2004). *Performance-Controlling professioneller Dienstleistungen: Gerichte im Effizienzvergleich*. München: Hampp.
- [So16] Song, M.-L., Fisher, R., Wang, J.-L., & Cui, L.-B. (2016). Environmental performance evaluation with big data: theories and methods. *Annals of Operations Research*.
- [SPS05] Sowlati, T., Paradi, J. C., & Suld, C. (2005). Information systems project prioritization using data envelopment analysis. *Mathematical and Computer Modelling*, 41(11–12), 1279–1298.
- [Th96] Thompson, R., Dharmapala, P., Diaz, J., Gonzales-Lima, M., & Thrall, R. (1996). DEA multiplier analytic center sensitivity with an illustrative application to independent oil companies. *Annals of Operations Research*, 66, 163–177.
- [VVL17] Varmaz, A., Varwig, A., & Laudi, P. (2017). Zentrale Ressourcenverteilung in der Gesamtbanksteuerung: Eine DEA-Anwendung. *Zeitschrift Für Bankrecht Und Bankwirtschaft*, 29(1).
- [VVP13] Varmaz, A., Varwig, A., & Poddig, T. (2013). Centralized resource planning and Yardstick competition. *Omega*, 41(1), 112–118.