prognos

Basel | Berlin | Bremen | Brüssel
Düsseldorf | Freiburg | München
Stuttgart

BH&W
Boos Hummel & Wegerich

Endfassung Teilbericht

Durchschnittsstrompreise

Forschungsvorhaben:
Weiterentwicklung des EEG 2014
im Hinblick auf die Kosten industrieller Verbraucher

Projekt 40/15

Studie im Auftrag des
Bundesministeriums für
Wirtschaft und Energie
(BMWi), Bonn / Berlin;
Projektträger Jülich

Ansprechpartner:
Marcus Koepp
Prognos AG
☎ 030 / 520 059 – 233

Berlin, 25.11.2016

Projektleitung:

Prognos AG
Europäisches Zentrum für Wirtschaftsforschung und Strategieberatung

Goethestraße 85
10623 Berlin
Telefon: 0 30 / 52 00 59 – 200
Fax: 030 / 52 00 59 – 201
www.prognos.com

Ansprechpartner:
Marcus Koepp
Telefon: 0 30 / 52 00 59 – 233
E-Mail: marcus.koepp@prognos.com

Mitarbeiter
Karsten Weinert

Unterauftragnehmer:

Boos Hummel & Wegerich (BH&W)

Zimmerstraße 56
10117 Berlin
Telefon: 0 30 / 20 09 547 – 0
Fax: 0 30 / 20 09 547 – 19
http://www.bhw-energie.de/

Ansprechpartner:
Dr. Heidrun Schalle, M.Jur.
Telefon: 0 30 / 20 09 547 – 14
E-Mail: schalle@bhw-energie.de

Die Arbeiten an diesem Bericht wurden Ende 2015 abgeschlossen.
Der Bericht berücksichtigt den zu diesem Zeitpunkt verfügbaren Datenbestand.

Inhalt

1 Ausgangslage und Zielsetzung 1

2 Beschreibung der Antragsdaten 2015 2

3 Basisdatensatz zur Preisermittlung 4

3.1 Begriffsbestimmungen 4

3.2 Plausibilitätstests 5

3.3 Aggregation der Daten auf Ebene der Abnahmestellen 6

3.3.1 Vollbenutzungsstunden 6

3.3.2 Netzebene des Strombezugs 7

4 Methoden zur Ermittlung der Durchschnittspreise 8

4.1 Quantile-basierte Methoden 8

4.1.1 Beschreibung 8

4.1.2 Sensitivitäten 9

4.2 Zweistufige Verfahren mit Clusterung und Klassifikation 10

4.2.1 Beschreibung 10

4.2.2 Sensitivitäten 11

4.3 Kriterien zur Bewertung der Methoden 12

5 Ausgewählte Ergebnisse und Bewertung 13

5.1 Überblick 13

5.2 Bevorzugte Berechnungsvariante 14

5.1 Sensitivitätsanalyse 16

5.1.1 Quantile-basierte Methoden 16

5.1.1 Zweistufiges Verfahren mit Clusterung / Klassifikation 19

6 Anhang 21

6.1 Beschreibung des VIK-Strompreisindex 21

6.2 Arbeitstreffen 22

6.3 Kennwerte der Verteilung der Merkmale im Antragsdatensatz zur Verwendung in der Simulation 23

Abbildungsverzeichnis

Abbildung 1:	Schematische Darstellung der Quantile-basierten Methoden	8
Abbildung 2:	Beispielhafte Ergebnisdarstellung des Quantile-basierten Verfahrens	9
Abbildung 3:	Beispielhafte Ergebnisdarstellung des zweistufigen Verfahrens	11
Abbildung 4:	Resultierendes Preistableau beim Quantile-basierten Verfahren mit 8x8 Gruppen bei fester Grenze bei 7000 Vollbenutzungsstunden (Stand der Daten: Anfang November 2015)	15
Abbildung 5:	Darstellung der dropouts, dropouts supercap sowie dropins supercap beim Quantile-basierten Verfahren mit 8x8 Gruppen bei fester Grenze bei 7000 Vollbenutzungsstunden (Stand der Daten: Anfang November 2015)	16
Abbildung 6:	Entwicklung der Genauigkeit (links) und dropouts (rechts) in Abhängigkeit der Gruppenanzahl bei der Quantile-basierten Methode mit fester 7.000 Vbh Grenze	18
Abbildung 7:	Entwicklung der Genauigkeit (jeweils links) und ökonomische Konsequenzen (jeweils rechts) für zwei Varianten des zweistufigen Verfahrens	20

Tabellenverzeichnis

Tabelle 1:	Beschreibung des Datensatzes der Abnahmestellen aus dem Antragsjahr 2015	2
Tabelle 2:	Beschreibung des Datensatzes der Unternehmen aus dem Antragsjahr 2015	3
Tabelle 3:	Plausibilitätstests	5
Tabelle 4:	Verteilung der Abnahmestellen bzgl. der Netzebene (nach Plausibilitätstest)	7
Tabelle 5:	Überblick über ausgewählte Berechnungsverfahren	14
Tabelle 6:	Sensitivität der Quantile-basierten Methode hinsichtlich der Bildung der Gruppengrenzen	18
Tabelle 7:	Sensitivität des zweistufigen Verfahrens (Clusterung / Klassifikation) hinsichtlich der einbezogenen Merkmale	19

1 Ausgangslage und Zielsetzung

Die im April 2014 veröffentlichten Umwelt- und Energiebeihilfeleitlinien der Europäischen Kommission (kurz: UEBLL) sehen vor, dass für die Gewährung von Ausnahmen von Finanzierungsbeiträgen zur Förderung erneuerbarer Energien gegenüber industriellen Stromverbrauchern nicht auf deren tatsächliche Stromkosten abgestellt werden darf. Stattdessen sollen die Stromkosten nach den einschlägigen Vorgaben in Annex 4 der UEBLL als Produkt des ggfs. standardisierten Stromverbrauchs und eines Durchschnittsstrompreises begriffen werden.

Im Rahmen des von Prognos und seinem Projektpartner für das BMWi vor kurzem abgeschlossenen Projekts „Transformation des Energiesystems […]“ (03MAP255) wurden heute vorhandene Datengrundlagen und noch bestehende Datenlücken für eine belastbare Berechnung der geforderten Durchschnittsstrompreise identifiziert. Verschiedene Verfahren zur Gruppierung der Antragsteller und zur daraus resultierenden Bildung von Durchschnittspreisen wurden in der Statistik-Software „R“ implementiert, auf die Ende 2014 verfügbaren Daten angewendet und die grundsätzlichen Auswirkungen verschiedener Methoden auf den heutigen Kreis der Antragsteller für die BesAR ermittelt. Allerdings zeigten die so erzielbaren Ergebnisse zum Teil noch unbefriedigend große Abweichungen von den individuellen Strompreisen der Antragsteller. Aufbauend auf diese Analysen wurden in enger Abstimmung mit dem BMWi erste Vorschläge für eine Weiterentwicklung der Datenerhebung durch das Bundesamt für Wirtschaft und Ausfuhrkontrolle (BAFA) entwickelt.

Der Schwerpunkt der Arbeiten in diesem Teilprojekt liegt deshalb einerseits auf der wissenschaftlichen Unterstützung des BMWi bei der mit den UEBLL konformen Weiterentwicklung und Umsetzung des Erhebungs- und Berechnungskonzepts für Durchschnittsstrompreise.

Der vorliegende Zwischenbericht stellt einen Arbeitsstand zum Ende November 2015 dar. In diesem Bericht werden potenziell geeignete Verfahren zur Berechnung von Durchschnittsstrompreisen aus den aktuellen Antragsdaten für die BesAR vorgestellt und anhand verschiedener Bewertungsverfahren getestet. Im Ergebnis wird eine von den Gutachtern bevorzugte Berechnungsvariante dargestellt. Für die endgültige Auswahl des optimalen Konzepts werden die erarbeiteten Varianten einer ökonomischen Folgenabschätzung unterzogen. Diese zielt darauf ab, positive und negative Effekte sowohl für die verschiedenen Gruppen von Antragstellern für die BesAR als auch bei den involvierten Vollzugsbehörden zu ermitteln.

2 Beschreibung der Antragsdaten 2015

Zur wissenschaftlichen Unterstützung bei der Weiterentwicklung und Umsetzung eines Erhebungs- und Berechnungskonzepts für die Durchschnittsstrompreise konnte auf die Antragsdaten zur Besonderen Ausgleichsregelung aus dem Jahr 2015 zurückgegriffen werden. Im Gegensatz zum Datensatz im Vorgängerprojekt zeichnet sich dieser Datensatz durch neue, zusätzliche Merkmale zum Lastprofil der Abnahmestellen aus.

Zum Zeitpunkt der Bearbeitung und der Dokumentation war die Prüfung und Überarbeitung des Datensatzes durch das Bundesamt für Wirtschaft und Ausfuhrkontrolle (BAFA) noch nicht abgeschlossen. Daher wurde mit einem vorläufigen Stand der Daten gearbeitet (**Stand: Anfang November 2015**).

Die folgenden beiden Tabellen dokumentieren die Struktur des Datensatzes. Die kursiv gedruckten Merkmale wurden bei der Bildung des Basisdatensatzes für die Preisermittlung (s. nächstes Kapitel) genutzt.

Tabelle 1: Beschreibung des Datensatzes der Abnahmestellen aus dem Antragsjahr 2015

Kurzname	Beschreibung	Datentyp / Einheit
vorgang	Identifikationsnummer Vorgang	Zeichenkette
unternehmen	*Identifikationsnummer Unternehmen*	*Zeichenkette*
wz	Wirtschaftszweig des Unternehmens	Zeichenkette
liste	Liste gemäß Anlage 4 zur § 64 EEG	Zeichenkette
spannungsebene	*Netzebene des Strombezugs: „n“ für Niederspannungsebene, „m“ für Mittelspannungsebene, „h0“ für Hochspannungsebene, „h1“ für Höchstspannungsebene*	*Zeichenkette*
hoechstlast.kw	*(gemessene) Jahreshöchstlast der Abnahmestelle in kW*	*Numerisch / kW*
mindestlast.kw	(gemessene) Jahresmindestlast der Abnahmestelle in kW	Numerisch / kW
durchschnittslast.kw	(gemessene) Jahresdurchschnittslast der Abnahmestelle in kW	Numerisch / kW
strommenge.gwh	*Strommenge, die über diese Abnahmestelle bezogen wurde in GWh*	*Numerisch / GWh*

Tabelle 2: Beschreibung des Datensatzes der Unternehmen aus dem Antragsjahr 2015

Kurzname	Beschreibung	Datentyp / Einheit
vorgang	Identifikationsnummer Vorgang	Zeichenkette
unternehmen	*Identifikationsnummer Unternehmen*	*Zeichenkette*
wz	Wirtschaftszweig des Unternehmens	Zeichenkette
liste	Liste gemäß Anlage 4 zu § 64 EEG	Zeichenkette
ski	*Stromkostenintensität*	*Numerisch / in Prozent, 100 = 100%*
beantragt.gwh	*zur Privilegierung beantragte Strommenge in GWh*	*Numerisch / GWh*
nichtbeantragt.gwh	*Stromverbrauch, für den keine Privilegierung beantragt wurde, in GWh*	*Numerisch / GWh*
weitergeleitet.gwh	*an Dritte weitergeleitete Strommengen in GWh*	*Numerisch / GWh*
bws.eur	*Bruttowertschöpfung des Unternehmens, in EUR*	*Numerisch / EUR*
umsatz.eur	Umsatz des Unternehmens in EUR	Numerisch / EUR
netzkosten.eur	Netznutzungsentgelte des Unternehmens (ohne die Kosten für etwaige Messstellenabrechnungen oder Konzessionsabgaben etc.), in EUR	Numerisch / EUR
netzerstattung.eur	rückerstattete Netznutzungsentgelte in EUR	Numerisch / EUR
stromkosten. mit_eeg.eur	*Stromkosten des Unternehmens einschließlich gezahlter und fiktiver EEG-Umlage, ohne Kosten für weitergeleiteten Strom, in EUR*	*Numerisch / EUR*
eeg.kosten.eur	*tatsächlich gezahlte EEG-Umlage des Unternehmens, in EUR*	*Numerisch / EUR*
eeg.kosten.fiktiv.eur	*fiktive, über die tatsächlich gezahlte EEG-Umlage hinausgehende Kosten für die EEG-Umlage, in EUR*	*Numerisch / EUR*
stromkosten. weiterleitung.eur	*Stromkosten für an Dritte weitergeleiteten Strom, in EUR*	*Numerisch / EUR*
ski.schwelle	*aufgrund der Liste gemäß Anlage 4 zu § 64 EEG geltende Schwelle für die Stromkostenintensität*	*Numerisch / in Prozent, 1=100%*

Aus Gründen des **Datenschutzes** war nur ein sehr eingeschränkter Zugang zu den Daten in den Räumlichkeiten des BMWi bei Anwesenheit eines Mitarbeiters des BMWi zulässig. Daher wurden zum Test der Implementierungen der Berechnungsverfahren und der zugehörigen Auswertung simulierte Daten verwendet, die eine ähnliche Struktur wie die tatsächlichen Daten haben. Die der Simulation zugrunde liegenden Parameter wurden aus den Daten ermittelt und sind in Anhang 6.3 dokumentiert.

3 Basisdatensatz zur Preisermittlung

3.1 Begriffsbestimmungen

Zentral für die Ermittlung der Durchschnittspreise sind die Stromkosten und die Strommenge der betrachteten Unternehmen. Eine genaue Bestimmung dieser Begriffe und eine Herleitung eines Berechnungsverfahrens für diese Begriffe ist für die weitere Untersuchung unerlässlich.

Mit dem Auftraggeber und dem BAFA wurden verschiedene Möglichkeiten der Begriffsbestimmung diskutiert und testweise als Verfahren implementiert.

Im Ergebnis wurden folgende Begriffsbestimmungen als zielführend ermittelt (die Variablen sind im vorigen Abschnitt näher beschrieben):

$$\begin{aligned} Strommenge.GWh \\ &:= beantragt.GWh + nichtbeantragt.GWh \\ &+ weitergeleitet.GWh \end{aligned}$$

$$\begin{aligned} Stromkosten.eur \\ &:= stromkosten.mit.eeg.eur - eeg.kosten.eur \\ &- eeg.kosten.fiktiv.eur \\ &+ stromkosten.weiterleitung.eur \\ &+ (Strommenge.GWh * 10^6 * 6{,}24/100) \end{aligned}$$

Bei der Strommenge werden sowohl zur Priviligierung beantragte als auch nicht beantragte Mengen berücksichtigt, denn darauf beziehen sich auch die angegebenen Stromkosten. Des Weiteren werden auch die an Dritte weitergeleiteten Strommengen einbezogen. Dies hat den Vorteil, dass keine Verzerrungen durch eine Über- oder Unterbewertung des durchgeleiteten Stroms auftreten kann.

Die nicht zur Priviligierung beantragte Strommenge hatte bei bisherigen Antragsverfahren keine entscheidende Bedeutung. Daher ist zu erwarten, dass in den Antragsdaten 2015 eine Reihe von Falscheinträgen mit fehlenden oder ungültigen Angaben vorzufinden sind. Mit den kommenden Antragsrunden wird sich die Datenqualität an dieser Stelle voraussichtlich weiter verbessern.

Gemäß § 64 Absatz 6 Nummer 2 und 3 EEG 2014 können sich die antragstellenden Unternehmen fiktiv so stellen, als wäre die EEG-Umlage im Nachweiszeitraum nicht durch die Besondere Ausgleichsregelung begrenzt gewesen. Die tatsächlich gezahlte EEG-Umlage sowie die Differenz bis zur fiktiven EEG-Umlage werden

jedoch separat angegeben. Für die an Dritte weitergeleitete Strommenge liegen keine Informationen zur EEG-Umlage vor, da dieser der Letztverbraucher trägt. Um die Stromkosten der Weiterleitung angemessen zu berücksichtigen, werden daher zunächst von den Stromkosten die sich auf die EEG-Umlage beziehenden (tatsächlichen und fiktiven) Kosten abgezogen. Hierauf werden die Kosten für den weitergeleiteten Strom sowie die fiktive EEG-Umlage in Höhe von 6,24 EUR-Cent je kWh auf die gesamte Strommenge aufaddiert.

Daten zur EEG-pflichtigen Eigenerzeugung liegen für das Antragsjahr 2015 zwar vor, wurden jedoch weder bei der Strommenge noch bei den Stromkosten berücksichtigt. Grund für dieses Vorgehen ist, dass die bei weitem meisten Eigenerzeugungsanlagen in diesem Antragsjahr nicht EEG-umlagepflichtig sind. Für kommende Antragsjahre ist die Berücksichtigung der Eigenerzeugung zu prüfen.

3.2 Plausibilitätstests

Zum Zeitpunkt der Bearbeitung und der Dokumentation war die Prüfung und Überarbeitung des Datensatzes durch das Bundesamt für Wirtschaft und Ausfuhrkontrolle (BAFA) noch nicht abgeschlossen. Es wurde mit einem vorläufigen Stand der Daten gearbeitet (**Stand: Anfang November 2015**).

Daher wurden vor der Bildung der Durchschnittspreise Plausibilitätstests durchgeführt, anhand derer offensichtlich fehlerbehaftete Daten von der weiteren Analyse ausgeschlossen wurden. Insgesamt wurden deshalb etwa 12 % der Daten nicht berücksichtigt. In Tabelle 3 sind die durchgeführten Plausibilitätstests beschrieben und die Anzahl der dabei herausfallenden Datensätze dokumentiert.

Tabelle 3: Plausibilitätstests

Kurzname	Beschreibung	Anzahl Beobachtungen
Strommenge	Strommenge fehlt oder ist kleiner als 10 MWh	1
Abnahmestelle	keine Abnahmestelle oder min. eine Abnahmestelle mit nicht-positiver Höchstlast oder Vollbenutzungsstunden kleiner 10 oder größer 8760	204
SKI	Stromkostenintensität auf Basis des individuellen Strompreises größer 0 oder kleiner dem listenspezifischen Schwellwert	23
Strompreis	Strompreis (ohne fiktive EEG-Umlage) kleiner 3 oder größer 20 EUR-Cent je kWh	26

Neben den Plausibilitätstests wurde auch nach auffälligen Werten mit Hilfe von statistischen Ausreißertests gesucht. Konkret wurden zwei Verfahren – zum einen mittels lokaler Ausreißerfaktoren

(LOF)[1], zum anderen mittels eines simultanen Cluster- und Ausreißerdetektionsverfahrens[2] – eingesetzt. Zwar konnten die Verfahren auffällige Punkte identifizieren, jedoch lieferten beide Verfahren teils unterschiedliche Ergebnisse. Zudem gibt es im Datensatz einige „auffällige“ Einträge, die nicht auf Fehleinträgen, sondern auf einer eben untypischen, aber dennoch plausiblen Verbrauchsstruktur beruhen. Aus diesem Grund wurden im weiteren Verlauf keine Datensätze aufgrund der statistischen Ausreißertests ausgeschlossen.

3.3 Aggregation der Daten auf Ebene der Abnahmestellen

3.3.1 Vollbenutzungsstunden

Auf Ebene der Abnahmestellen liegen für das Antragsjahr 2015 erstmalig Informationen u. a. zur Höchstlast vor. Aus diesem Merkmal werden als neue Variable zur Erklärung des Strompreis die Vollbenutzungsstunden als Quotient von bezogener Strommenge und Höchstlast gebildet.

Um diese neue Variable bei der Bildung von Durchschnittspreisen mit berücksichtigen zu können, ist eine Aggregation auf der Ebene der Unternehmen erforderlich. Hierzu wurden mit dem Auftraggeber und dem BAFA verschiedene Möglichkeiten diskutiert und testweise als Verfahren implementiert:

1) Vollbenutzungsstunden der Abnahmestelle mit dem höchsten Strombezug,
2) Vollbenutzungsstunden der Abnahmestelle mit der höchsten Netzebene des Strombezugs,
3) strommengengewichtetes Mittel der Vollbenutzungsstunden aller Abnahmestellen des Unternehmens,
4) Vollbenutzungsstunden aus Gesamtstrommenge und Mittelwert von höchstmöglicher Gesamtlast aller Abnahmestellen und größter Einzel-Jahreshöchstlast.

Im Ergebnis wurde das dritte Verfahren, „strommengengewichtetes Mittel der Vollbenutzungsstunden aller Abnahmestellen“ ausgewählt. Gegen Verfahren 1) und 2) spricht, dass bewusst auf verfügbare Informationen verzichtet wird. Verfahren 4) ist nicht sachgerecht, da es zu Vollbenutzungsstunden größer als 8.760 Stunden führen kann.[3] Die Abweichungen insbesondere hinsichtlich

---

[1] M. M. Breunig, Hans-Peter Kriegel, R.T. Ng, J. Sander: LOF: Identifying Density-based Local Outliers. In: ACM SIGMOD Record. Nr. 29, 2000, S. 93

[2] Sanjay Chawla ∗ Aristides Gionis: k-means--: A unified approach to clustering and outlier detection. 2013

[3] Beispiel: Abnahmestelle 1: 1 MW bei 7.500 MWh, Abnahmestelle 2: 250 kW bei 2.500 MWh führt zu 10.000 MWh / 0,5*(1,25 MW + 1 MW) = 8.889 Vbh

der ökonomischen Konsequenzen sind im Übrigen robust gegenüber der Wahl des Verfahrens.

3.3.2 Netzebene des Strombezugs

Auf Ebene der Abnahmestelle liegt für das Antragsjahr 2015 erstmalig die Netzebene des Strombezugs vor. Die Netzebene hat einen Bezug zu den Netznutzungsentgelten und kann daher potenziell den Strompreis erklären.

Die Möglichkeiten zur Berücksichtigung der Netzebene wurden mit dem Auftraggeber und dem BAFA diskutiert und testweise implementiert. Im Ergebnis zeigte sich, dass der Großteil der Abnahmestellen an die Mittelspannungsebene (88 %) angeschlossen ist (s. Tabelle 4). Eine Differenzierung nach diesem Merkmal führt daher schnell zu methodisch schwierig handhabbaren geringen Feldbelegungen für die anderen Netzebenen. Insgesamt ist der Erklärungsbeitrag eher gering, weshalb im weiteren Verlauf der Untersuchung die Netzebene nicht weiter berücksichtigt wurde.

Tabelle 4: Verteilung der Abnahmestellen bzgl. der Netzebene (nach Plausibilitätstest)

Netzebene	**Abnahmestellen auf dieser Netzebene**	
	Anzahl	**Prozent**
Niederspannung	86	3,3%
Mittelspannung	2304	88,4%
Hochspannung	198	7,6%
Höchstspannung	19	0,7%
Abnahmestellen insgesamt	**2607**	**100%**

4 Methoden zur Ermittlung der Durchschnittspreise

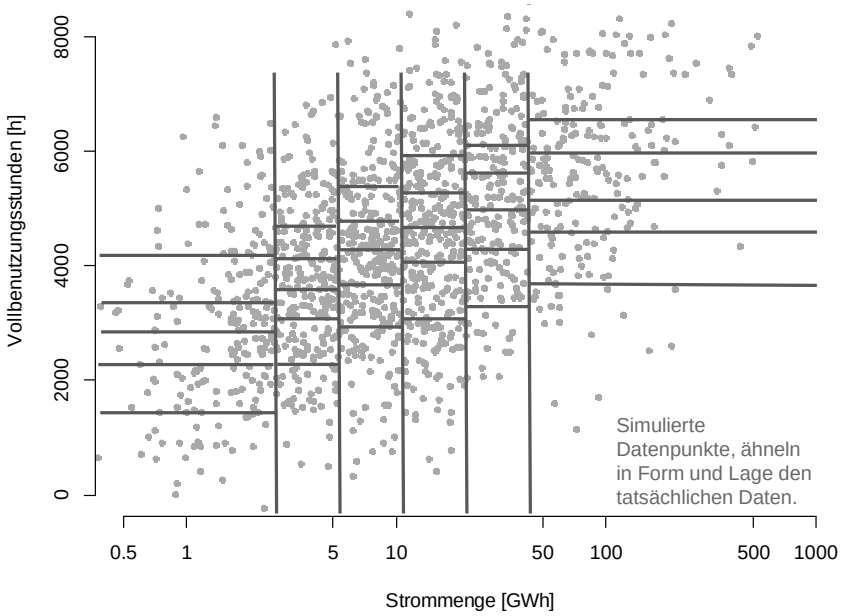
Zur Ermittlung der Durchschnittspreise wurden die Ansätze aus dem kürzlich abgeschlossenen Projekts „Transformation des Energiesystems […]“ (03MAP255) aufgegriffen und auf die neu verfügbare Datensituation angewendet. Im Folgenden werden die grundsätzlichen Ansätze und die betrachteten Sensitivitäten erläutert. Anschließend werden die zur Bewertung herangezogenen Kriterien dargestellt.

4.1 Quantile-basierte Methoden

4.1.1 Beschreibung

Bei den Quantile-basierten Methoden werden die Beobachtungen zunächst hinsichtlich der Strombezugsmenge auf eine festgelegte Anzahl von Gruppen aufgeteilt, so dass auf jede Gruppe die gleiche Anzahl von Beobachtungen entfällt. In Abbildung 1 entspricht dieser Schritt dem Einziehen der senkrechten Linien. Im zweiten Schritt werden innerhalb jedes Strombezugsmengenbands eine Aufteilung hinsichtlich der Vollbenutzungsstunden vorgenommen, wiederum derart, dass auf jede Gruppe die gleiche Anzahl von Beobachtungen entfällt. Dieser Schritt ist in der Abbildung durch die jeweils waagerechten Liniensegmente schematisch dargestellt.

Abbildung 1: Schematische Darstellung der Quantile-basierten Methoden



Das Ergebnis dieser Methode ist eine Preistabelle, die in Abhängigkeit von Strombezugsmenge und Vollbenutzungsstunden einen anzulegenden Durchschnittspreis ausweist (siehe Abbildung 2 für eine schematische Darstellung).

Abbildung 2: Beispielhafte Ergebnisdarstellung des Quantile-basierten Verfahrens

Stromverbrauch	Vollbenutzungsstunden							
0 – 2,7 GWh	1286 VBh	1995VBh	2608VBh	3095VBh	3587 VBh	4455 VBh		
	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	
2,7 GWh – 4,2 GWh	2203 VBh	2788 VBh	3491 VBh	4011 VBh	4518 VBh	5051 VBh		
	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	
4,2 GWh – 6,3 GWh	2491 VBh	3360 VBh	3900 VBh	4280 VBh	4705 VBh	5284 VBh		
	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	
6,3 GWh – 9,6 GWh	2728 VBh	3764 VBh	4292 VBh	4806 VBh	5177 VBh	5699 VBh	7000 VBh	
	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh
9,6 GWh – 14,8 GWh	3140 VBh	4055 VBh	4591 VBh	4907 VBh	5408 VBh	6059 VBh	7000 VBh	
	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh
14,8 GWh – 25,7 GWh	3229 VBh	4134 VBh	4742 VBh	5157 VBh	5616 VBh	6068 VBh	7000 VBh	
	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh
25,7 GWh – 54,0 GWh	3212 VBh	4106 VBh	4774 VBh	5305 VBh	5784 VBh	6215 VBh	7000 VBh	
	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh
Größer als 54 GWh	3810 VBh	4731 VBh	5245 VBh	5643 VBh	6001 VBh	6453 VBh	7000 VBh	
	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh	___ ct/kWh

Der Vorteil dieses Verfahrens ist, dass mittels einer einheitlichen Systematik annähernd gleich große Gruppen gebildet werden.

Der potenzielle Nachteil dieses Verfahrens ist, dass für die Preisbildung relevante Aspekte, wie z. B. die 7.000 h-Regel gemäß § 19 Abs. 2 StromNEV, nicht berücksichtigt werden. Ein weiterer Nachteil ist, dass aufgrund der rechtsschiefen Verteilung der Strombezugsmenge (die x-Achse in Abbildung 1 ist logarithmisch dargestellt) in den Gruppen des größten Strombezugsmengenbands eine deutliche Variabilität des Strombezugs zu beobachten ist (z. B. von 54 GWh bis über 3,7 TWh bei 8x8 Gruppen).

4.1.2 Sensitivitäten

Zur Adressierung der potenziellen Nachteile wurde die Quantile-basierte Methode in verschiedene Richtungen variiert:

1) **7.000 h Band**: In dieser Variante wird nach dem ersten Schritt (Bildung der Strombezugmengenbänder) eine feste Gruppe von Beobachtungen mit mehr als 7.000 Vbh eingeführt. Die verbleibenden Beobachtungen mit weniger als 7.000 Vbh werden danach hinsichtlich der Vollbenutzungsstunden in gleich große Gruppen aufgeteilt. Sofern in der Gruppe über 7.000 Vbh nur wenige Beobachtungen auftreten (weniger als 20), wird für dieses Strombezugsmengenband die ursprüngliche ausschließlich Quantile-basierte Methode eingesetzt.

2) **Oberstes Strombezugsmengenband teilen**: Hier wird das Strombezugsmengenband mit den größten Strommengen nochmals in zwei gleichgroße Gruppen aufgeteilt.

Um bei den Varianten 1) und 2) eine Vergleichbarkeit mit der Basisvariante zu ermöglichen, sind die Verfahren so gewählt, dass die Gesamtzahl der gebildeten Gruppen bei den Varianten gleichbleibt.

3) **Gruppengrößen**: Als weitere Variante wird die Zahl der gebildeten Klassen variiert. Es werden alle Varianten von 1 bis 20 Vollbenutzungsstundenklassen bei 1 bis 20 **Strombezugsmengenbändern** implementiert. Die Gruppengrößen 6x6, 8x8 und 10x10 sind dabei insofern vertieft betrachtet worden, dass dort auch jeweils die Varianten 1) und 2) untersucht wurden.

4.2 Zweistufige Verfahren mit Clusterung und Klassifikation

4.2.1 Beschreibung

Bei dem zweistufigen Verfahren mit Clusterung und Klassifikation wird im Unterschied zum Quantile-basierten Verfahren keine geordnete Struktur (aufsteigende Strombezugsmengen, aufsteigenden Vollbenutzungsstundenklassen) zugrunde gelegt. Stattdessen wird in einem ersten Schritt ein statistischer Algorithmus (sog. kmeans[4]) herangezogen, der auf Basis der vorliegenden Merkmale[5] einschließlich des Strompreises möglichst ähnliche Gruppen bildet. Diese Gruppen müssen nicht einer rechteckigen Struktur wie beim Quantile-basierten Verfahren folgen, nicht-lineare Strukturen (sog. Cluster) sind denkbar und die Regel. Lediglich die Anzahl der Gruppen wird von außen vorgegeben.

Im zweiten Schritt werden auf der Basis von einem oder mehreren anderen Merkmalen (z. B. Strommenge und Vollbenutzungsstunden) die Unternehmen diesen gebildeten Preisgruppen zugeordnet. Hierzu wird ein rekursives Partitionierungsverfahren verwendet[6]. Ergebnis dieses Schrittes ist ein sog. Klassifikationsbaum, mit dem die Unternehmen einzelnen Gruppen zugeordnet werden. Der Baum besteht aus Ästen und Verzweigungen. An jeder Verzweigung wird eine bestimmte Bedingung geprüft (z. B. *Strommenge >= x ?* oder *Vollbenutzungsstunden < y ?*) und eine Zuordnung vorgenommen, ohne dabei die individuellen Strompreise einzubeziehen. Beispielhaft ist ein Klassifikationsbaum als mögliches Ergebnis des Verfahrens in Abbildung 3 dargestellt.

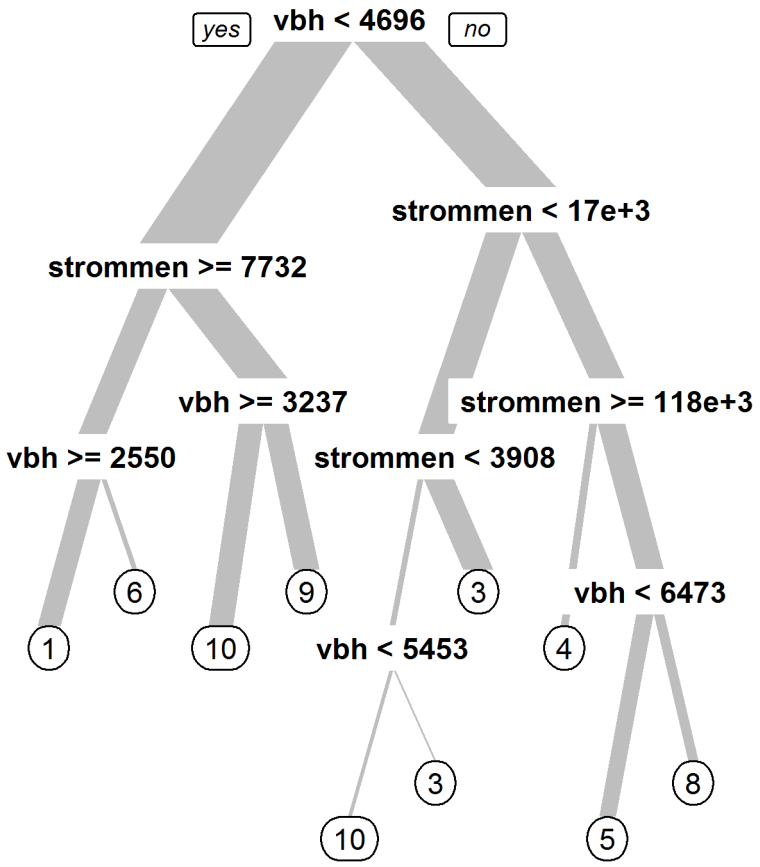
---

4 Backhaus et al. (2010): Multivariate Analysemethoden: Eine anwendungsorientierte Einführung, Springer, Berlin

5 Die Merkmale werden jeweils z-transformiert (Mittelwert=0, Standardabweichung=1)

6 Loh (2011): Classification and regression trees, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Volume 1, Issue 1, pages 14–23, January/February 2011

Abbildung 3: Beispielhafte Ergebnisdarstellung des zweistufigen Verfahrens



Der Vorteil dieser Methode liegt darin, dass keine Struktur den Daten von außen vorgegeben wird, sondern vielmehr vom Algorithmus „entdeckt“ wird. Damit kann das Verfahren Besonderheiten im Datensatz potenziell besser darstellen.

Der Nachteil des Verfahrens liegt zum einen darin, dass im Allgemeinen keine Optimalität des Verfahrens garantiert werden kann und das Ergebnis somit von den Startwerten beeinflusst ist. Zum anderen kann es im zweiten Schritt des Verfahrens zu Fehlklassifikationen kommen, da in diesem Schritt die Preisinformation nicht mehr verwendet wird. Die Anzahl der Fehlklassifikationen, also der verfahrensinternen Inkonsistenz steigt tendenziell mit der Zahl der im ersten Schritt gebildeten Gruppen.

4.2.2 Sensitivitäten

Zur genaueren Untersuchung des Verfahrens wurden folgende Varianten betrachtet:

1) **Anzahl Cluster**: Die Anzahl der Gruppen, die im ersten Schritt des Verfahrens gebildet werden, wird hier von 4 bis 60 variiert.

2) **Inklusion von Merkmalen**: Hierbei werden die verfügbaren Merkmale – Strommenge, Vollbenutzungsstunden, Handelsintensität, Bruttowertschöpfung – in unterschiedlichen Kombinationen bei den beiden Schritten des Verfahrens einbezogen und mit dem Referenzverfahren (Strommenge und Vollbenutzungsstunden) verglichen.

3) **Alternative Methode mit verbesserter Wahl der Startwerte**: Hier wird eine erweiterte Variante des kmeans-Algorithmus angewendet, dessen Anwendung in der Theorie bessere Ergebnisse verspricht[7] und so die Abhängigkeit vom Startwert reduzieren kann.

4.3 Kriterien zur Bewertung der Methoden

Zur Bewertung der Methoden zur Durchschnittspreisermittlung wurden zwei Kriterien angelegt:

- **Genauigkeit der berechneten Durchschnittspreise**: Abweichung zwischen zugeordnetem Durchschnittspreis und individuell gemeldeten Preisen
- **Ökonomische Konsequenzen**: Der Anteil der Unternehmen, die bei Anwendung der Durchschnittspreise die SKI-Schwelle nicht mehr erreichen (*dropouts*), das *supercap*[8] verlieren oder neu unter das *supercap* fallen (*dropin*)

Zum Zeitpunkt der Bearbeitung und der Dokumentation war die Prüfung und Überarbeitung des Datensatzes durch das Bundesamt für Wirtschaft und Ausfuhrkontrolle (BAFA) noch nicht abgeschlossen. Es wurde mit einem vorläufigen Stand der Daten gearbeitet (**Stand: Anfang November 2015**).

Daher ist zu erwarten, dass ein Teil der *dropouts* auf Fehleingaben bzw. fehlenden Angaben beruht und durch die weitere Datenprüfung und -aufbereitung bereinigt werden. Insgesamt ist die Zahl der *dropouts* durch die nachfolgende Betrachtung tendenziell überschätzt.

Zudem ist bei der Betrachtung der Antragsdaten 2015 nicht feststellbar, ob und wieviele weitere Unternehmen durch die Anwendung von Durchschnittspreisen neu zusätzlich erfolgreich eine Priviligierung beantragen werden.

---

7 Arthur, D. and S. Vassilvitskii (2007). k-means++: The advantages of careful seeding. In H. Gabow (Ed.), Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms [SODA07], Philadelphia, pp. 1027-1035. Society for Industrial and Applied Mathematics

8 Unter *supercap* wird im Folgenden die Regelung nach § 64 (2) Nr. 3. und Nr. 4. EEG 2014 verstanden, nach der die zu zahlende EEG-Umlage für besonders stromkostenintensive Unternehmen auf weniger als 15 % der jeweils gültigen Höhe, aber mindestens auf 0,05 Cent/kWh bzw. 0,1 Cent/kWh begrenzt wird.

5 Ausgewählte Ergebnisse und Bewertung

In diesem Abschnitt werden zunächst ausgewählte Varianten der Berechnungsverfahren eingeordnet und hinsichtlich der Kriterien aus Kapitel 4.3 bewertet. Daraus wird eine bevorzugte Berechnungsvariante abgeleitet und detailliert dargestellt. Ein kursorischer Überbick über die Ergebnisse der Sensitivitätsrechnungen bildet den Abschluss des Kapitels. Die im Folgenden dargestellten Ergebnisse beruhen auf dem durch das Bundesamt für Wirtschaft und Ausfuhrkontrolle (BAFA) noch nicht abschließend geprüften und überarbeiteten Datensatz (**Stand: Anfang November 2015**).

5.1 Überblick

Tabelle 5 zeigt die Genauigkeit und und die ökonomischen Konsequenzen für vier ausgewählte Berechnungsverfahren:

- Quantile-basiertes Verfahren mit 6x6 Gruppen mit fester Grenze bei 7.000 Vollbenutzungsstunden,
- Quantile-basiertes Verfahren mit 8x8 Gruppen mit fester Grenze bei 7.000 Vollbenutzungsstunden,
- Quantile-basiertes Verfahren mit 10x10 Gruppen mit fester Grenze bei 7.000 Vollbenutzungsstunden,
- zweistufiges Verfahren mit Clusterung und Klassifikation mit 10 Gruppen und Strommenge sowie Vollbenutzungsstunden als berücksichtigte Merkmale.

Zur Beurteilung der Genauigkeit des Verfahrens ist die Abweichung zwischen gemeldeten und Durchschnittspreisen – nach oben oder unten – entscheidend. Daher wird in der Tabelle 5 der Anteil der Unternehmen ausgewiesen, deren Abweichung weniger als 10 %, zwischen 10 und 20 %, zwischen 20 und 50 % oder mehr als 50 % beträgt. Je genauer ein Verfahren, desto höher ist der Anteil in der Zeile „kleiner 10 %“.

Zur Beurteilung der ökonomischen Konsequenzen des Verfahrens wird der Anteil der Unternehmen und der Anteil an der zur Priviligierung beantragten Strommenge ausgewiesen, die bei Anlegung des Durchschnittspreises die Priviligierung der BesAR verlieren würden (*dropouts*). Außerdem wird der Anteil der Unternehmen dargestellt, die das *supercap* verlieren (*dropouts supercap*) bzw. gewinnen (*dropins supercap*) würden.

Zu erkennen ist in Tabelle 5, dass die ausgewählten Quantile-basierten Verfahren sich hinsichtlich Genauigkeit und ökonomischer Konsequenzen in sehr ähnlicher Größenordnung befinden. Beim Übergang zu 10x10 Gruppen wird die Genauigkeit geringfügig erhöht, jedoch erhöht sich gleichzeitig die Zahl der *dropouts*.

Tabelle 5: Überblick über ausgewählte Berechnungsverfahren

	Quantile 6x6 7.000 Vbh	**Quantile 8x8 7.000 Vbh**	**Quantile 10x10 7.000 Vbh**	**Zweistufig: Clusterung/ Klassifikation**
Genauigkeit des Verfahrens: Anteil der Unternehmen mit einer Abweichung der gemeldeten Preise von den Gruppenpreisen um…				
<10%	78,26%	78,15%	79,05%	76,73%
10-20%	16,41%	16,52%	16,09%	17,84%
20-50%	5,07%	5,07%	4,59%	5,17%
>50%	0,26%	0,26%	0,26%	0,26%
Ökonomische Konsequenzen				
Dropouts [Unternehmen]	3,64%	3,48%	3,69%	4,22%
Dropouts [Strommenge]	2,09%	1,95%	1,91%	2,17%
Dropouts supercap [Unternehmen]	2,90%	2,85%	2,90%	2,80%
Dropins supercap [Unternehmen]	1,74%	1,85%	1,58%	1,53%

Das zweistufige Verfahren mit Clusterung und Klassifikation schneidet hinsichtlich Genauigkeit und Anteil der Unternehmen, die die Priviligierung verlieren, sichtbar schlechter ab. Jedoch scheinen zumindest teilweise andere Unternehmen mit im Durchschnitt geringerem Stromverbrauch unter den *dropouts* zu sein, denn der Anteil der Strommenge, die aus der Priviligierung fällt, ist geringer als bei den anderen Verfahren.

5.2 Bevorzugte Berechnungsvariante

Unter Berücksichtigung der Tatsache, dass der Datenstand vorläufigen Charakter hat, empfehlen die Gutachter als bevorzugte Berechnungsvariante die Verwendung von 8x8 Quantilen mit fester Gruppengrenze bei 7.000 Vollbenutzungsstunden. Wenn dabei in einer Gruppe über 7.000 Vbh weniger als 20 Beobachtungen vorliegen, sollte für dieses Strombezugsmengenband die Quantile-basierte Methode ohne die 7.000 Vbh-Grenze eingesetzt werden.

Der Stromverbrauch ist die zentrale Einflussgröße für den Strompreis und muss – auch nach Vorgabe der EU-Kommission – berücksichtigt werden. Die Kombination von Stromverbrauch und Vollbenutzungsstunden hat hohen Erklärungsgehalt für die Strombezugskosten.

Die Differenzierung in rund 60 Ausprägungen (8x8) bildet ein großes Spektrum der Abnahmefälle ab. Eine Vergrößerung der Matrix bringt nur noch geringe Verbesserungen.

Über die Quantile-Methode gebildete Matrizen weisen weniger *dropouts* auf als die multivariate Clusterung. Zudem ergibt die Quantile-basierte Methode ein übersichtliches Preistableau (Abbildung 4), welches für Unternehmen einfach zu handhaben ist.

Die Berücksichtigung der 7.000 Vbh-Grenze verringert (wenn auch nur geringfügig) die *dropouts* gegenüber der ausschließlichen Quantile-Methode. Fast in jeder Gruppe sind *dropouts* zu finden, so dass eine Benachteiligung für bestimmte Abnahmefälle nicht zu beoachten ist. Des Weiteren ist zu erwarten, dass sich nach abschließender Prüfung und Überarbeitung der Antragsdaten durch das BAFA die Anzahl der *dropouts* eher reduzieren wird.

Abbildung 4 zeigt das resultierende Preistableau für das empfohlene Verfahren. Hierin ist die EEG-Umlage enthalten. Zu erkennen ist die Tendenz zu niedrigeren Preisen bei zunehmender Strombezugsmenge und / oder zunehmenden Vollbenutzungsstunden.

Abbildung 4: Resultierendes Preistableau beim Quantile-basierten Verfahren mit 8x8 Gruppen bei fester Grenze bei 7.000 Vollbenutzungsstunden (Stand der Daten: Anfang November 2015)

Stromverbrauch	Vollbenutzungsstunden							
	1313 VBh	1938 VBh	2420 VBh	2871 VBh	3306 VBh	3771 VBh	4617 VBh	
0 – 2,63 GWh	**16,12 ct/kWh**	**16,74 ct/kWh**	**16,6 ct/kWh**	**16,23 ct/kWh**	**15,54 ct/kWh**	**15,6 ct/kWh**	**15,53 ct/kWh**	**14,97 ct/kWh**
	2133 VBh	2697 VBh	3193 VBh	3663 VBh	4159 VBh	4696 VBh	5168 VBh	
2,63 GWh – 4,17 GWh	**16,03 ct/kWh**	**16,53 ct/kWh**	**15,86 ct/kWh**	**15,43 ct/kWh**	**15,03 ct/kWh**	**14,68 ct/kWh**	**14,82 ct/kWh**	**14,45 ct/kWh**
	2365 VBh	3236 VBh	3771 VBh	4104 VBh	4370 VBh	4916 VBh	5558 VBh	
4,17 GWh – 6,27 GWh	**15,36 ct/kWh**	**15,57 ct/kWh**	**15,28 ct/kWh**	**15,18 ct/kWh**	**14,90 ct/kWh**	**14,84 ct/kWh**	**14,35 ct/kWh**	**14,64 ct/kWh**
	2820 VBh	3771 VBh	4302 VBh	4755 VBh	5103 VBh	5489 VBh	6277 VBh	
6,27 GWh – 9,47 GWh	**15,3 ct/kWh**	**15,36 ct/kWh**	**14,88 ct/kWh**	**14,62 ct/kWh**	**14,51 ct/kWh**	**13,87 ct/kWh**	**14,12 ct/kWh**	**14,64 ct/kWh**
	2883 VBh	3882 VBh	4486 VBh	4797 VBh	5144 VBh	5712 VBh	6266 VBh	
9,47 GWh – 14,48 GWh	**15,36 ct/kWh**	**14,74 ct/kWh**	**14,72 ct/kWh**	**14,63 ct/kWh**	**14,12 ct/kWh**	**13,96 ct/kWh**	**13,81 ct/kWh**	**13,69 ct/kWh**
	3154 VBh	3990 VBh	4726 VBh	5188 VBh	5643 VBh	6073 VBh	7000 VBh	
14,48 GWh–25,24 GWh	**14,58 ct/kWh**	**14,44 ct/kWh**	**14,3 ct/kWh**	**13,81 ct/kWh**	**13,93 ct/kWh**	**14,1 ct/kWh**	**13,26 ct/kWh**	**12,37 ct/kWh**
	3212 VBh	4078 VBh	4713 VBh	5195 VBh	5723 VBh	6167 VBh	7000 VBh	
25,24 GWh – 52,8 GWh	**13,64 ct/kWh**	**14,1 ct/kWh**	**14,0 ct/kWh**	**13,73 ct/kWh**	**13,65 ct/kWh**	**13,45 ct/kWh**	**13,05 ct/kWh**	**12,33 ct/kWh**
	3785 VBh	4554 VBh	5143 VBh	5616 VBh	6005 VBh	6439 VBh	7000 VBh	
Größer als 52,8 GWh	**13,82 ct/kWh**	**13,29 ct/kWh**	**13,11 ct/kWh**	**12,67 ct/kWh**	**12,52 ct/kWh**	**12,99 ct/kWh**	**11,99 ct/kWh**	**11,68 ct/kWh**

Abbildung 5: Darstellung der dropouts, dropouts supercap sowie dropins supercap beim Quantile-basierten Verfahren mit 8x8 Gruppen bei fester Grenze bei 7.000 Vollbenutzungsstunden (Stand der Daten: Anfang November 2015)

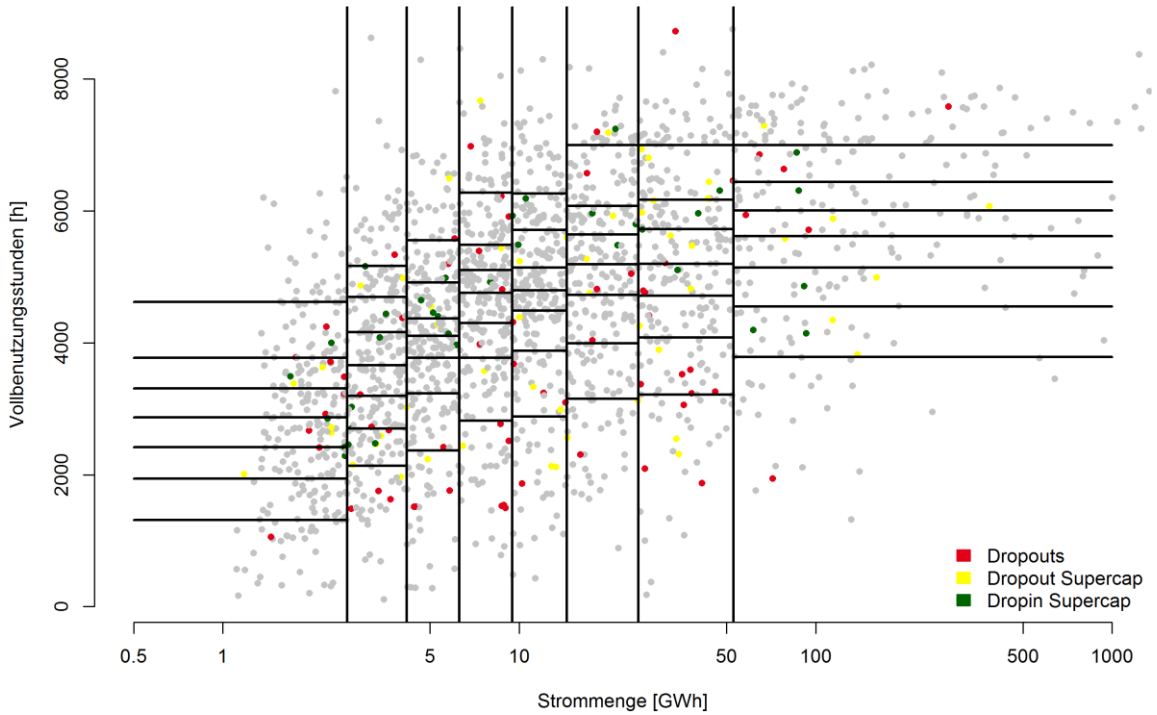


Abbildung 5 veranschaulicht die ökonomischen Wirkungen bei Anwendung dieses Verfahrens zur Durchschnittspreisbildung. Zu erkennen ist, dass die *dropouts* in nahezu allen Gruppen zu finden sind. Eine tendenzielle Benachteiligung einzelner Größenklassen ist nicht beobachten.

5.1 Sensitivitätsanalyse

5.1.1 Quantile-basierte Methoden

Bei den Quantile-basierten Methoden wurden zwei Varianten bei der Gruppenbildung geprüft. Zusammen mit der ausschließlich auf Quantilen basierenden Gruppenbildungsmethode ergeben sich somit drei Verfahren, die bei Betrachtung einer gleich großen Anzahl von Gruppen (z. B. 8x8, 8 Strombezugsmengenbänder, 8 Gruppen nach Vollbenutzungsstunden je Strombezugsmengenband) untereinander vergleichbar sind.

1) **Quantile 8x8**: Dies ist das ausschließlich auf Quantilen basierende Verfahren. Im ersten Schritt werden die Beobachtungen hinsichtlich der Strombezugsmenge auf eine festgelegte Anzahl von Gruppen aufgeteilt, so dass auf jede Gruppe die gleiche Anzahl von Beobachtungen entfällt. Im zweiten Schritt wird innerhalb jedes Strombezugsmengenbands eine Aufteilung

hinsichtlich der Vollbenutzungsstunden vorgenommen, wiederum derart, dass auf jede Gruppe die gleiche Anzahl von Beobachtungen entfällt.

2) **Quantile 8x8, feste Grenze bei 7.000 Vbh**: In dieser Variante wird nach dem ersten Schritt (Bildung der Strombezugmengenbänder) eine feste Gruppe von Beobachtungen mit mehr als 7.000 Vbh eingeführt. Die verbleibenden Beobachtungen mit weniger als 7.000 Vbh werden danach hinsichtlich der Vollbenutzungsstunden in sieben gleich große Gruppen aufgeteilt. Sofern in der Gruppe über 7.000 Vbh nur wenige Beobachtungen auftreten (weniger als 20), werden die Gruppengrenzen für dieses Strombezugsmengenband ausschließlich mittels acht Quantilen ohne die 7.000 Vbh-Grenze bestimmt.

3) **Quantile 8x8, oberstes Strombezugsmengenband zweigeteilt:** Hier wird im ersten Schritt das Strombezugsmengenband in sieben Quantile unterteilt. Das Quantil mit der größten Strommenge wird nochmals in zwei gleich große Gruppen aufgeteilt. Die Aufteilung hinsichtlich der Vollbenutzungsstunden erfolgt wiederum mit fester 7.000 Vbh Grenze (s. Variante 2).

Tabelle 6 stellt die drei Varianten und ihre Ergebnisse für die Bewertungskriterien dar. Insbesondere bei der Genauigkeit sind die Unterschiede gering. Eine zusätzliche Aufteilung des Strombezugsmengenbandes mit den größten Mengen reduziert zwar die *dropouts* hinsichtlich der Strommenge leicht, die Zahl der Unternehmen, die das *supercap* verlieren, erhöht sich jedoch leicht.

Als weitere Sensitivität wurde der Einfluss der Anzahl der Gruppen untersucht. Hierzu wurde das Verfahren mit fester Grenze bei 7.000 Vbh für eine von 1 bis 20 variierende Anzahl von Strombezugsmengenbändern und Vbh-Kategorien untersucht. Abbildung 6 zeigt die dabei beobachtete Genauigkeit (Anteil der Unternehmen mit weniger als 10 % Abweichung von Gruppenpreisen zu Individualpreisen, links) bzw. ökonomischen Konsequenzen (*dropout*-Anteil: Anteil der Unternehmen, die die Priviligierung verlieren würden, rechts). Zu erkennen ist, dass sich mit steigender Gruppenzahl die Genauigkeit erhöht und der Anteil der *dropouts* reduziert. Nennenswerte Effekte treten jedoch erst bei recht hoher Gruppenzahl auf. Eine Schwelle mit einem qualitativem Sprung ist nicht auszumachen, insbesondere bei den *dropouts* gibt es offenbar neben der Gruppenanzahl weitere, das Ergebnis beeinflussende Variablen.

Tabelle 6: Sensitivität der Quantile-basierten Methode hinsichtlich der Bildung der Gruppengrenzen

	Quantile 8x8	**Quantile 8x8 feste Grenze bei 7.000 Vbh**	**Quantile 8x8, oberstes Strombezugsmengenband zweigeteilt, feste Grenze bei 7.000 Vbh**
Genauigkeit des Verfahrens: Anteil der Unternehmen mit einer Abweichung der gemeldeten Preise von den Gruppenpreisen um…			
<10%	78,89%	78,15%	78,47%
10-20%	17,04%	16,52%	16,46%
20-50%	4,85%	5,07%	4,80%
>50%	0,26%	0,26%	0,26%
Ökonomische Konsequenzen			
Dropouts [Unternehmen]	3,48%	3,48%	3,48%
Dropouts [Strommenge]	2,05%	1,95%	1,74%
Dropouts supercap [Unternehmen]	2,85%	2,85%	3,11%
Dropins supercap [Unternehmen]	1,74%	1,85%	1,85%

Abbildung 6: Genauigkeit (links) und dropouts (rechts) in Abhängigkeit der Gruppenanzahl bei der Quantile-basierten Methode mit fester 7.000 Vbh Grenze

Anteil Unternehmen, in denen die Abweichung Individualpreis zu Gruppenpreis kleiner als 10% ist

61,5	66,8	70,6	72,6	72,6	72,9	73,2	72,8	73,0	73,3	72,8	72,8	73,0	72,7	73,2	72,4	73,2	72,6	72,6	73,0																
65,5	73,6	74,2	75,2	74,6	75,1	74,6	75,2	75,1	75,5	75,4	75,6	75,2	75,0	75,6	75,2	75,5	75,6	75,8	75,9																
69,1	73,4	75,2	74,6	75,9	76,1	76,0	75,9	76,2	76,3	75,7	76,2	76,1	76,3	76,0	76,5	76,7	76,4	76,5	76,9																
70,4	74,0	76,2	76,2	76,1	76,2	76,7	76,3	76,5	76,2	76,0	76,7	76,3	76,7	76,5	77,2	76,4	76,4	76,2	76,4																
70,8	75,2	76,8	77,2	77,0	76,7	76,5	76,8	77,1	77,2	77,3	77,7	77,1	77,6	77,8	76,8	77,2	76,9	78,1	77,5																
71,0	76,0	77,1	77,2	77,5	77,1	77,0	77,7	77,9	78,4	77,9	77,9	78,2	77,7	77,6	77,4	78,0	77,8	78,1	77,9																
71,8	75,7	76,2	76,8	77,2	77,4	77,2	77,2	77,7	77,4	77,9	78,0	78,2	77,3	77,6	77,8	77,2	77,7	77,4	77,9																
71,7	76,1	77,1	77,5	77,3	76,8	77,4	77,2	77,4	77,7	77,8	78,2	78,0	78,0	78,0	78,3	77,0	78,6	78,6	78,7																
71,8	76,1	77,5	77,1	77,7	77,3	77,7	77,9	78,0	77,7	77,3	77,7	78,0	78,6	78,3	78,0	78,0	78,4	78,4	78,7																
70,6	76,3	76,5	77,3	77,8	77,0	77,4	77,7	78,0	77,9	78,0	77,1	77,9	78,1	78,1	78,3	78,1	78,0	78,1	78,3																
71,8	76,5	77,5	77,4	77,4	77,6	77,1	77,5	78,2	77,9	77,9	78,0	77,8	78,3	78,6	78,1	78,4	78,1	78,9	78,1																
71,8	76,5	77,5	77,4	77,6	78,0	78,3	78,2	78,2	78,6	77,9	78,4	78,4	78,6	78,7	78,7	78,4	78,9	79,1	79,3																
71,8	76,7	77,4	77,5	77,9	77,7	78,2	77,4	78,1	77,9	77,9	77,3	77,7	77,9	79,1	78,9	79,1	78,4	79,0	79,4																
71,5	75,2	77,1	76,5	77,3	76,9	77,4	77,1	77,6	77,7	77,7	76,8	77,9	78,8	78,7	79,2	78,4	79,0	78,2	78,8																
71,5	76,4	77,3	77,5	77,8	77,8	77,8	77,7	78,1	78,4	78,3	78,4	78,9	79,5	79,9	79,5	79,3	79,5	79,0	80,2																
71,7	76,6	76,9	76,7	77,9	77,5	78,0	77,3	77,8	78,6	77,7	78,2	78,2	78,7	79,0	78,6	78,8	79,8	79,2	79,2																
72,2	75,8	77,4	77,9	78,0	77,7	78,2	77,6	77,9	78,3	78,2	78,6	78,0	78,9	78,7	78,4	78,1	79,2	78,7	80,2																
72,2	75,6	77,6	77,0	78,1	77,7	77,9	78,2	78,4	78,1	79,0	78,8	78,8	79,3	79,2	79,7	79,7	80,4	80,0	80,1																
72,5	76,0	77,4	77,5	77,3	77,7	78,6	77,7	78,4	78,9	78,8	79,6	79,3	79,7	80,3	80,0	80,2	80,4	80,4	80,8																
72,2	75,4	77,5	77,9	78,1	77,7	78,4	78,8	79,0	79,0	78,9	79,7	79,2	79,2	79,1	80,0	79,6	79,8	81,0	80,8																

Anzahl Mengenbänder ↓

Anzahl Vbh-Kategorien →

Anteil der Unternehmen, die die Priviligierung bei Ansatz der Gruppenpreise verlieren würden (*dropouts*)

6,71	5,76	5,23	5,02	5,12	5,12	5,18	4,97	5,07	5,18	4,70	4,81	4,91	4,75	4,97	4,81	4,81	4,65	4,81	4,81																
5,60	4,86	4,81	4,86	4,86	4,86	4,54	4,49	4,75	4,70	4,75	4,81	4,65	4,60	4,54	4,81	4,70	4,70	4,44	4,60																
5,28	4,65	4,33	4,33	4,07	4,01	4,17	4,23	4,28	4,01	4,07	3,96	4,17	4,28	4,23	3,80	3,96	4,07	4,12	4,33																
5,44	4,75	4,33	4,49	4,07	4,17	4,01	4,28	4,12	4,12	4,23	4,28	4,33	4,07	4,07	4,12	4,23	4,01	4,23	3,91																
5,39	4,38	4,23	3,86	4,28	4,01	4,07	4,07	4,12	4,12	3,96	3,80	3,96	3,86	3,91	4,17	4,07	3,96	4,12	3,80																
5,18	4,33	4,28	4,01	4,01	3,96	4,12	3,80	4,01	3,70	3,96	3,70	3,86	4,07	4,07	4,23	3,91	4,17	4,12	3,91																
4,86	4,38	4,44	4,07	4,01	3,86	3,96	4,01	4,17	4,07	4,01	3,96	4,07	3,91	3,96	3,80	4,07	3,59	3,70	3,70																
5,18	4,12	4,44	4,07	4,07	3,91	4,23	3,91	4,23	4,17	4,01	3,96	4,01	4,17	4,23	3,91	4,07	3,91	4,17	4,01																
5,23	4,17	4,23	4,17	4,12	3,86	4,12	3,96	4,17	4,28	4,23	4,12	4,01	4,17	4,12	4,12	3,91	4,12	4,07	4,12																
5,12	4,44	4,01	4,17	4,17	3,91	3,96	3,96	4,07	3,86	3,80	3,96	3,75	3,91	3,91	3,86	4,12	3,86	3,86	3,65																
5,28	4,23	4,33	4,23	4,54	4,17	4,12	3,96	4,07	4,23	4,28	4,12	4,01	3,96	4,07	3,75	3,75	3,80	3,70	3,86																
5,28	4,12	4,17	4,12	4,17	4,07	4,23	3,80	3,91	4,44	4,17	4,23	4,17	4,28	3,91	4,07	4,01	3,86	3,91	4,01																
5,12	4,33	4,17	3,91	4,01	3,80	4,17	3,86	4,07	3,96	4,49	4,01	4,07	4,17	4,01	4,07	4,17	3,91	3,86	3,96																
5,12	3,96	4,28	4,28	4,01	3,75	3,91	3,80	3,49	3,86	3,59	3,80	3,86	3,86	3,91	3,80	3,75	3,91	3,59	3,75																
5,18	4,23	4,38	4,17	4,23	4,01	3,96	4,28	4,23	3,91	4,12	4,23	4,17	3,86	3,86	4,01	3,91	3,96	4,17	3,86																
5,18	4,38	4,44	3,91	4,23	4,07	3,96	4,17	4,12	4,07	4,23	4,01	3,91	3,91	3,91	4,07	3,86	3,49	3,59	3,86																
5,18	4,17	4,38	4,07	4,49	4,07	3,96	4,12	4,23	3,96	4,12	4,07	4,28	3,91	3,70	3,75	3,65	3,70	3,70	3,96																
4,75	4,07	4,28	4,01	4,28	4,23	4,12	4,23	4,17	4,38	4,33	4,33	3,86	4,01	3,54	4,07	4,01	3,70	3,96	3,65																
5,12	4,12	4,28	4,01	4,23	3,96	4,49	4,17	4,12	4,07	4,33	3,96	3,80	3,86	3,59	4,17	4,17	3,65	3,91	3,96																
4,91	4,38	4,49	4,17	3,96	3,70	3,91	3,96	3,65	3,96	4,23	3,49	3,70	4,01	3,70	3,91	3,65	3,49	3,49	3,43																

Anzahl Vbh-Kategorien →

5.1.1 Zweistufiges Verfahren mit Clusterung / Klassifikation

Als Sensitivität beim zweistufigen Verfahren wurden die verfügbaren Merkmale in zwei Kombinationen untersucht. Die Ergebnisse hinsichtlich Genauigkeit und ökonomischer Konsequenzen sind in Tabelle 7 dargestellt. Zu erkennen ist, dass die Berücksichtigung der Bruttowertschöpfung als weitere erklärende Variable über die Strombezugsmenge und Vollbenutzungsstunden hinaus zu keiner nennenswerten Verbesserung der Genauigkeit oder Verringerung der *dropouts* führt, sondern im Gegenteil das Ergebnis verschlechtert. Eine mögliche Erklärung hierfür ist, dass die Bruttowertschöpfung mit der Strombezugsmenge korrelieren dürfte und dadurch zu einer Verzerrung bei der Gruppenbildung führt.

Tabelle 7: Sensitivität des zweistufigen Verfahrens (Clusterung / Klassifikation) hinsichtlich der einbezogenen Merkmale

	Zweistufig, Strommenge, Vbh	**Zweistufig, Strommenge, Vbh, BWS**
Genauigkeit des Verfahrens: Anteil der Unternehmen mit einer Abweichung der gemeldeten Preise von den Gruppenpreisen um…		
<10%	76,73%	74,83%
10-20%	17,84%	18,94%
20-50%	5,17%	5,96%
>50%	0,26%	0,26%
Ökonomische Konsequenzen		
Dropouts [Unternehmen]	4,22%	4,27%
Dropouts [Strommenge]	2,17%	2,37%
Dropouts supercap [Unternehmen]	2,80%	2,69%
Dropins supercap [Unternehmen]	1,53%	1,74%

Verwendete Abkürzungen: Vbh: Vollbenutzungsstunden, BWS: Bruttowertschöpfung

Als weitere Sensitivität wurde der Einfluss der Verfahrensparameter Clusteranzahl und Methode zur Wahl der Startwerte untersucht. Die Ergebnisse sind in Abbildung 7 dargestellt. Entgegen der Intuition schneidet das klassische kmeans-Verfahren mit zufälliger Startwertwahl besser ab als das „kmeans++"-Verfahren, bei dem die Startwerte auf Basis der bereits gewählten Clusterzentren und deren Lage in den Daten gewählt werden (siehe die im Kap. 4.2 zitierte Veröffentlichung für eine ausführliche Beschreibung). Allerdings zeigt das klassische kmeans-Verfahren deutlich höhere Schwankungen hinsichtlich der Genauigkeit und *dropouts*. Das in der Startwertwahl modifizierte Verfahren zeigt eine mit der Clusteranzahl steigende Genauigkeit und einen rückläufigen Anteil der *dropouts*, zumindest bis zu einer Clusteranzahl von etwa 30 bis 40. Danach sinkt die Qualität der Verfahren für diese beiden Bewertungskriterien. Dies ist möglicherweise auf die mit zu hoher Gruppenzahl eintretende höhere Fehlklassifikation in der zweiten Stufe zu erklären.

Abbildung 7: Entwicklung der Genauigkeit und der ökonomischen Konsequenzen (dropouts) für zwei Varianten des zweistufigen Verfahrens

	kmeans++			klass. kmeans	
	Genauigkeit	dropouts		Genauigkeit	dropouts
3	73,4	4,27	3	74,1	4,54
6	74,5	3,91	6	74,8	4,17
9	75,5	4,27	9	75,7	4,43
12	76,5	3,96	12	75,7	3,64
15	76,1	3,96	15	76,8	4,06
18	77,2	3,85	18	77,2	3,69
21	77,4	3,85	21	78,3	3,59
24	77,5	4,12	24	77,7	3,91
27	76,9	3,64	27	77,0	3,75
30	77,5	3,69	30	77,2	3,91
33	76,9	3,59	33	77,6	3,75
36	77,6	3,59	36	78,4	4,01
39	78,4	3,91	39	78,0	3,80
42	77,0	3,85	42	77,2	3,85
45	78,2	3,75	45	75,9	4,06
48	77,0	3,69	48	77,9	3,69
51	77,0	3,80	51	77,2	4,12
54	77,6	3,64	54	77,2	3,96
57	77,8	3,75	57	77,4	3,80
60	77,3	3,96	60	77,7	3,91

Genauigkeit= Anteil der Unternehmen, in denen die Abweichung zwischen Individual- und Gruppenpreis mehr als 10 % beträgt.
dropouts= Anteil der Unternehmen, die die Priviligierung bei Ansatz der Gruppenpreise verlieren würden
kmeans++: Verfahren mit modifizierter Startwertbestimmung
klass. kmeans: „klassisches" kmeans-Verfahren mit zufälliger Startwertbestimmung

6 Anhang

6.1 Beschreibung des VIK-Strompreisindex

Da die Durchschnittspreise erst im folgenden Antragsjahr zur Anwendung kommen, besteht prinzipiell das Risiko, dass sich innerhalb eines Jahres einzelne Strompreiskomponenten verändern und sich somit auch der Gesamtpreis deutlich ändert. Denkbar ist dies beispielsweise beim Börsenpreis oder den Netznutzungsentgelten.

Ein Ansatz, diese Preisbewegungen bei der BesAR mit zu berücksichtigen, ist die Verwendung oder Nachbildung eines Strompreisindexes wie z. B. des VIK Strompreisindexes, der im Folgenden näher dargestellt wird.

Seit 2002 veröffentlicht der VIK monatlich einen VIK-Strompreisindex für Mittelspannungskunden in Industrie und Gewerbe. Dieser VIK-Strompreisindex basiert einerseits auf den Strompreisen der European Energy Exchange (EEX) und andererseits auf den Netznutzungsentgelten sechs großer deutscher Netzbetreiber. Ab November 2014 wurde dieser VIK-Index in VIK-Basisindex umbenannt und weitergeführt. Gleichzeitig wurde ein zusätzlicher VIK-Index mit der Bezeichnung VIK-Endpreisindex neu eingeführt, welcher auch staatlich induzierte Steuern, Abgaben und Umlagen beinhaltet.

Als Energiepreiskomponente liegen beiden VIK-Indizes die monatlichen Durchschnittspreise des Vormonats am EEX-Terminmarkt für die kommenden vier noch handelbaren Quartalsprodukte zugrunde. Base- und Peakloadanteile werden in Abhängigkeit von typischen Jahresbenutzungsstunden für Industriekunden (3.000, 4.000, 5.000 und 6.000 h/a) gewichtet. Mit steigenden Jahresbenutzungsstunden steigt dabei auch der Baseloadanteil - bspw. für 6.000 h/a auf 92 %.

Der zweite gemeinsame Bestandteil beider VIK-Indizes sind die Netznutzungsentgelte der sechs Verteilnetzbetreiber Westnetz GmbH, Bayernwerk AG, Netze BW GmbH, MITNETZ STROM GmbH, Stromnetz Hamburg GmbH und Stromnetz Berlin GmbH für die Entnahme aus der Mittelspannung (Netzebene 5) - entsprechend den o. g. Benutzungsstunden. Die Summe der durchschnittlichen Energie- und Netzpreise aller Regelzonen bildet somit die Basis für die monatlich veröffentlichten VIK-Indizes. Bezugsmonat für die Indizierung ist der Januar 2002.

Der VIK-Endpreisindex beinhaltet zusätzlich zu den o. g. Energie- und Netzkomponenten des VIK-Basisindex noch von Industrieun-

ternehmen üblicherweise zu zahlende gesetzliche Steuern, Abgaben und Umlagen. Bei der Festlegung der teilweise individuellen Belastungshöhen wurden beispielsweise für das Jahr 2014 folgende Annahmen getroffen, um eine hohe Repräsentanz zu erreichen:

Die Konzessionsabgabe (KA) für Sondervertragskunden floss mit 1,1 €/MWh in den VIK-Endpreisindex ein. Die KWK-G-Umlage, die §19 Abs.2 StromNEV-Umlage sowie die §17f EnWG-Offshore-Umlage wurden entsprechend als reduzierten Umlage der Letztverbraucherkategorie B (LV Kat. B) mit 0,5 €/MWh angenommen. Die Umlage für abschaltbare Lasten nach §18 AbLaV wurde im Jahr 2014 mit 0,09 €/MWh berücksichtigt. Der Anteil der Stromsteuer ist entsprechend dem Stromsteuerspitzenausgleich unternehmensindividuell. Für den VIK-Endpreisindex wurde im Jahr 2014 ein Selbstbehalt von 20 % (4,306 €/MWh) angenommen. Die EEG-Umlage wurde in voller Höhe (62,40 €/MWh - 2014) berücksichtigt, da über die Hälfte des von der Industrie verbrauchten Stroms - insbesondere aus der Mittelspannung - nicht unter die besondere Ausgleichsregelung gem. §64 EEG fällt und somit nicht reduziert war.

Mit diesen Annahmen spiegelt der VIK-Endpreisindex eine realistische Preisentwicklung inklusive der staatlichen Steuern und Abgaben wider und bietet Informationen zur Gesamtkostenentwicklung für typische Mittelspannungsstromkunden.

6.2 Arbeitstreffen

Aus Gründen des Datenschutzes war nur ein sehr eingeschränkter Zugang zu den Daten in den Räumlichkeiten des BMWi bei Anwesenheit eines Mitarbeiters des BMWi zulässig. Daher fanden eine Reihe von Arbeitstreffen statt, wo vorbereitete Auswertungen auf den Datensatz angewendet und die Ergebnisse diskutiert wurden.

- 13.8. interner Kick-Off des Projektes bei Prognos
- 26.8. im BMWi (Workshop zusammen mit dem BAFA)
- Arbeitstreffen am 8.9., 24.9., sowie am 6.10., 8.10., 16.10., 28.10. und 17.11. im BMWi
- 30.10. Stakeholder-Workshop im BMWi

6.3 Kennwerte der Verteilung der Merkmale im Antragsdatensatz zur Verwendung in der Simulation

Zur Programmierung der Auswertungen wurden aus Datenschutzgründen simulierte Daten verwendet, die in Lage und Struktur den tatsächlichen Daten ähneln und mit folgendem Programmcode generiert wurden[9]:

```
sim_data_as <- function(...) {
  n_unt <- 1900
  n_as <- 2600
  unternehmen <- sample(as.character(seq(1000,3000)), n_unt, replace=FALSE)
  anz_as <- sample(c(
    rep(1, round(n_as*0.83)),
    rep(2, round(n_as*0.1)),
    rep(3, round(n_as*0.04)),
    rep(4, round(n_as*0.025)),
    rep(5, round(n_as*0.005))
  ), n_unt-1, replace=TRUE)
  anz_as <- c(anz_as, n_as-sum(anz_as))

  ret <- data.frame(
    unternehmen= rep(unternehmen, times=anz_as),
    strommenge.mwh=exp(rnorm(n_as, mean=9, sd=1.35)),
    spannungsebene=sample(c("h1", rep("h0",8), rep("m",88), rep("n",3)), n_as, replace=TRUE),
    stringsAsFactors=FALSE
  )
  ret[, "strommenge.gwh"] <- ret[, "strommenge.mwh"] / 1000
  ret[, "vbh"] <- 2850+745*log(ret[, "strommenge.gwh"])+rnorm(n=n_as, mean=0, sd=1470)
  ret[, "hoechstlast.kw"] <- ret[, "strommenge.mwh"] * 1000 / ret[, "vbh"]

  return(ret)
}

sim_data_unt <- function(data_as, ...) {
  ret <- data.frame(unternehmen=unique(data_as[, "unternehmen"]), stringsAsFactors=FALSE)
  n <- nrow(ret)
  sp <- split(data_as, data_as$unternehmen)
  ret[,"ski.schwelle"] <- sample(
    c(rep(0,51), rep(0.14, 250), rep(0.17, 1139), rep(0.2, 422)),
    n, replace=TRUE
  )
  ret[, "ski.errechnet"] <- ret[,"ski.schwelle"] + exp(rnorm(n, mean=-2.35, sd=1.477))
  menge_agg <- function(dset) sum(dset[, "strommenge.gwh"])
  ret[, "strommenge.gwh"] <- sapply(sp, FUN=menge_agg)[ret[, "unternehmen"]]
  vbh_agg <- function(dset)
    as.numeric(dset[, "vbh"] %*% dset[, "strommenge.mwh"] / sum(dset[, "strommenge.mwh"]))
  ret[, "vbh"] <- sapply(sp, FUN=vbh_agg)[ret[, "unternehmen"]]
  ret[, "strompreis"] <- 10.8-0.0003088*ret[,"vbh"]+exp(-0.57101*ret[, "strommenge.gwh"])+
    rnorm(n, mean=0, sd=1.5)
  return(ret)
}
```

Die weiteren Merkmale wurden aus den hier dargestellten simulierten Daten analog zu Kapitel 3 ermittelt und dem Datensatz hinzugefügt.

[9] Als Programmiersprache wurde R verwendet.