
Fraunhofer
Innovation Technologie Wissen Menschen
Fraunhofer
ITWM

FINANZMATHEMATISCHE MODELLIERUNG AM SEKUNDÄRREGELLEISTUNGSMARKT

Strommarkttreffen, 30. Juni 2017

Prilly Oktoviany

prilly.oktoviany@itwm.fraunhofer.de

in Zusammenarbeit mit
Dr. Andreas Wagner
Lukas Müller

Fraunhofer
ITWM

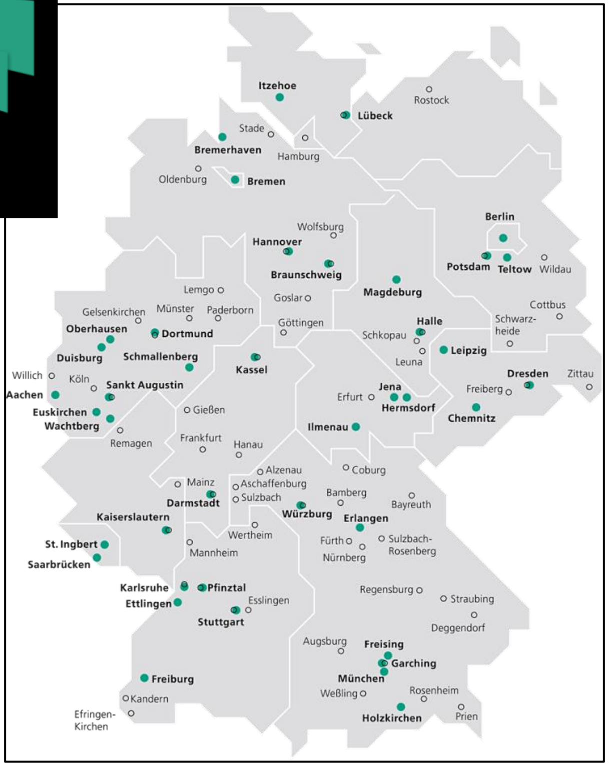
Abteilung »Finanzmathematik« am Fraunhofer ITWM

Energy Finance ist ein Schwerpunkt der Abteilung

- **Fraunhofer-Gesellschaft** zur Förderung der angewandten Forschung e.V.
  - Fraunhofer: 67 Institute mit ca. 24.000 Mitarbeitern
  - Finanzierung: 70 % Industrie und öffentliche Aufträge
- **Fraunhofer ITWM** (Kaiserslautern)
  - Weltgrößtes Institut für Industriemathematik
  - 240 Mitarbeiterinnen und Mitarbeiter in 9 Abteilungen
- **Abteilung »Finanzmathematik«**
  - Hochqualifiziertes Team mit 13 Finanzmathematikern
  - Wissenschaftliche Berater sichern Bezug zu aktueller Forschung
  - Zahlreiche Projekte mit Industriepartnern
- **Software-Produkte der Abteilung**
  - ALMSim (Finanzmarktmodellierung für Versicherungen)
  - Commodity Risk Manager (Risikomanagement Energiewirtschaft)
  - Auffälligkeitsdetektion (Betrugsdetektion Abrechnungsdaten)
  - Structured Products Pricer (Bewertung Zinsprodukte)

Software-Lösungen

- Mathematik für die **Finanzwirtschaft**
- Data-Science für das **Controlling**
- Mathematik für die **Energiewirtschaft**



Fraunhofer ITWM

Die Modellierung der historischen Ausschreibungswochen stellt die Grundlage für die Simulation zukünftiger Ausschreibungswochen dar

Ziel: Modellierung historischer Ausschreibungswochen in den Gebotskomponenten Leistungspreis, angebotene Leistung und Arbeitspreis am Sekundärregelleistungsmarkt

Ausgangssituation: In einer Pay-As-Bid Auktion ist die eigene Positionierung und Gebotsabgabe für die eigenen potentiellen Umsätze relevant

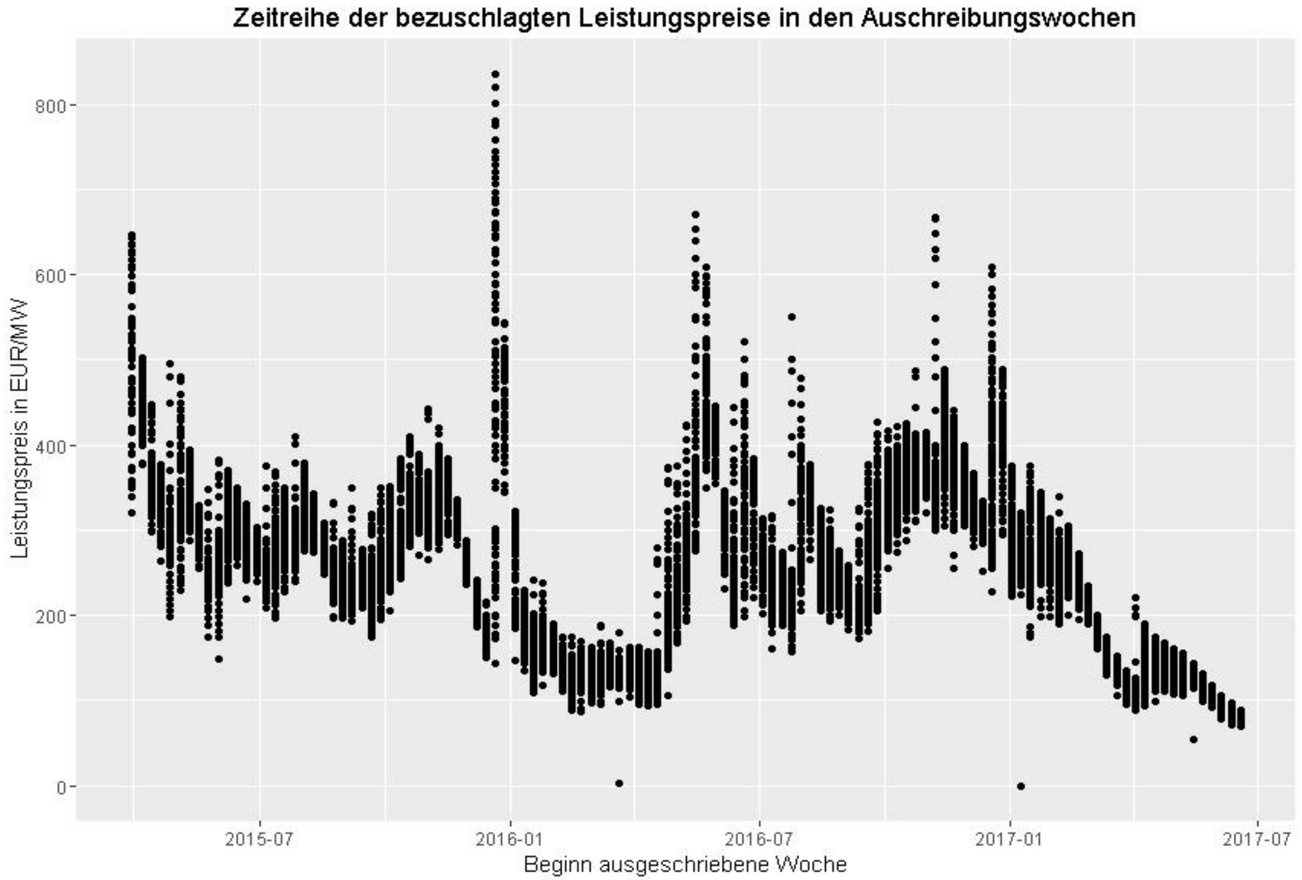
Datengrundlage: Wöchentliche Auktionsergebnisse werden auf regelleistung.net veröffentlicht, jedoch nur die akzeptierten Gebote, d.h. nur alle Gebote bis zum Grenzleistungspreis

Vorgehensweise:

- Entwicklung statistischer Modelle für die einzelnen Gebotskomponenten unter Verwendung historischer Daten und unter Berücksichtigung der Unvollständigkeit der Daten
- Grundannahme für die Einzelmodellierung: Unabhängigkeit der Gebotskomponenten

Fraunhofer ITWM

Die Leistungspreise der historischen Ausschreibungswochen für POS_HT zeigen irreguläre Verläufe und starke Streuungen

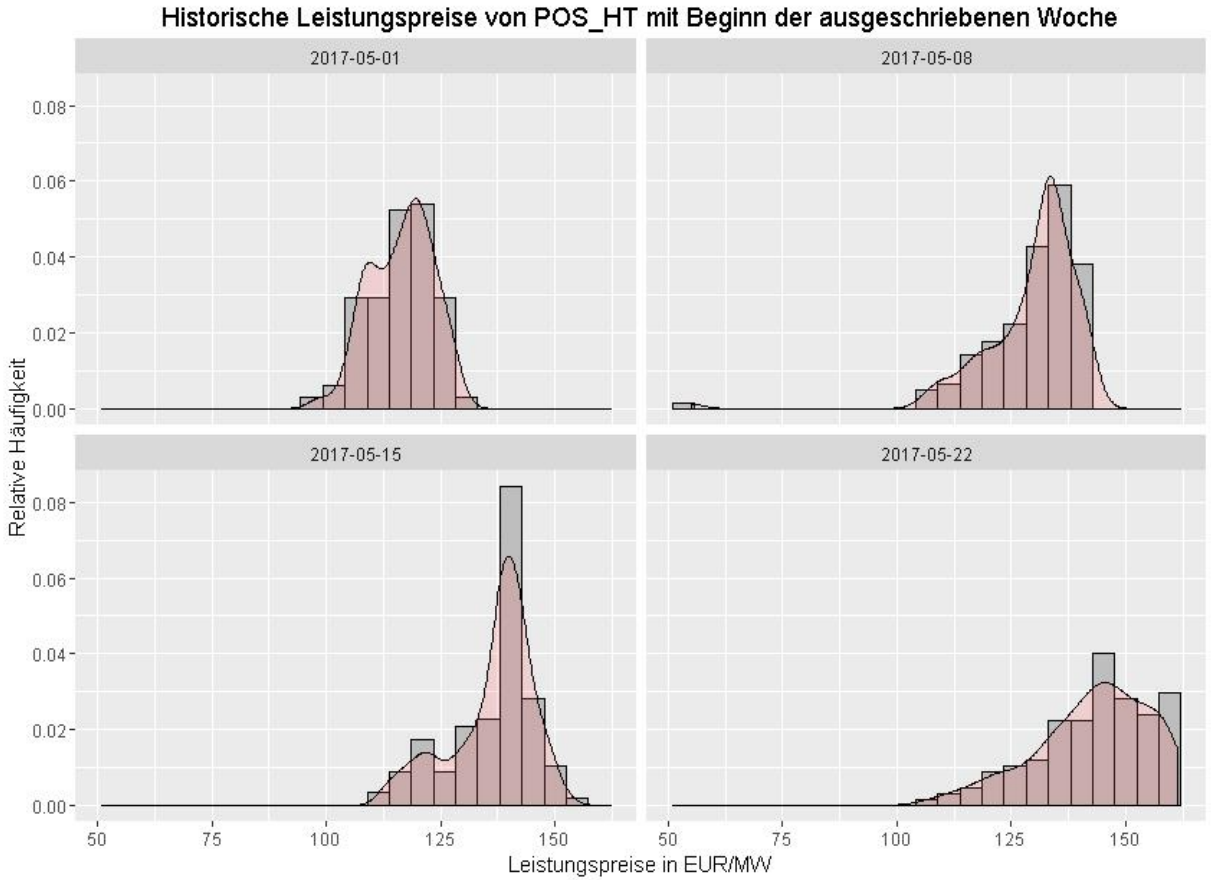


- Durch Senkung der Mindestgröße ein sinkender Trend der Preise für den aktuellen Zeitraum erkennbar
- Schwankungen zwischen den Wochen und starke Streuung der Gebote innerhalb der Wochen erschweren eine Zeitreihenanalyse

Können trotz der irregulären Verläufe der Gebotsstruktur die Wochen mittels einer gemeinsamen Methode modelliert werden?

Fraunhofer
ITWM

Die Unvollständigkeit der Daten und Clusterbildungen müssen in der Modellierung der Leistungspreise berücksichtigt werden



Die historischen, akzeptierten Leistungspreise sind um einen bestimmten Preis zentriert, in manchen Wochen auch um mehrere

- Zwei (einfache) Verteilungsannahmen unter Berücksichtigung unvollständiger Daten
  - Unimodaler Ansatz: **Trunkierte Normalverteilung**
  - Multimodaler Ansatz: **Gaussche Mischverteilung auf abgeschnittene Daten**

Aufgrund der Unvollständigkeit der Daten greifen wir auf die Modellierung mit der trunkierten Normalverteilung zurück

Eine trunkierte Normalverteilung ist die Wahrscheinlichkeitsverteilung einer normalverteilten Zufallsvariable, die von unten und/oder oben beschränkt ist

- In unserem Fall: $\text{Leistungspreis} \coloneqq X \sim N(\mu, \sigma^2)$ hat bedingt auf $0 < X < Grenzleistungspreis\ (GLP)$ eine trunkierte Normalverteilung mit Dichtefunktion

$$f(x|\mu, \sigma, 0, GLP) = 1_{(0,GLP)} \frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\sigma\left(\Phi\left(\frac{GLP-\mu}{\sigma}\right) - \Phi\left(\frac{0-\mu}{\sigma}\right)\right)}$$

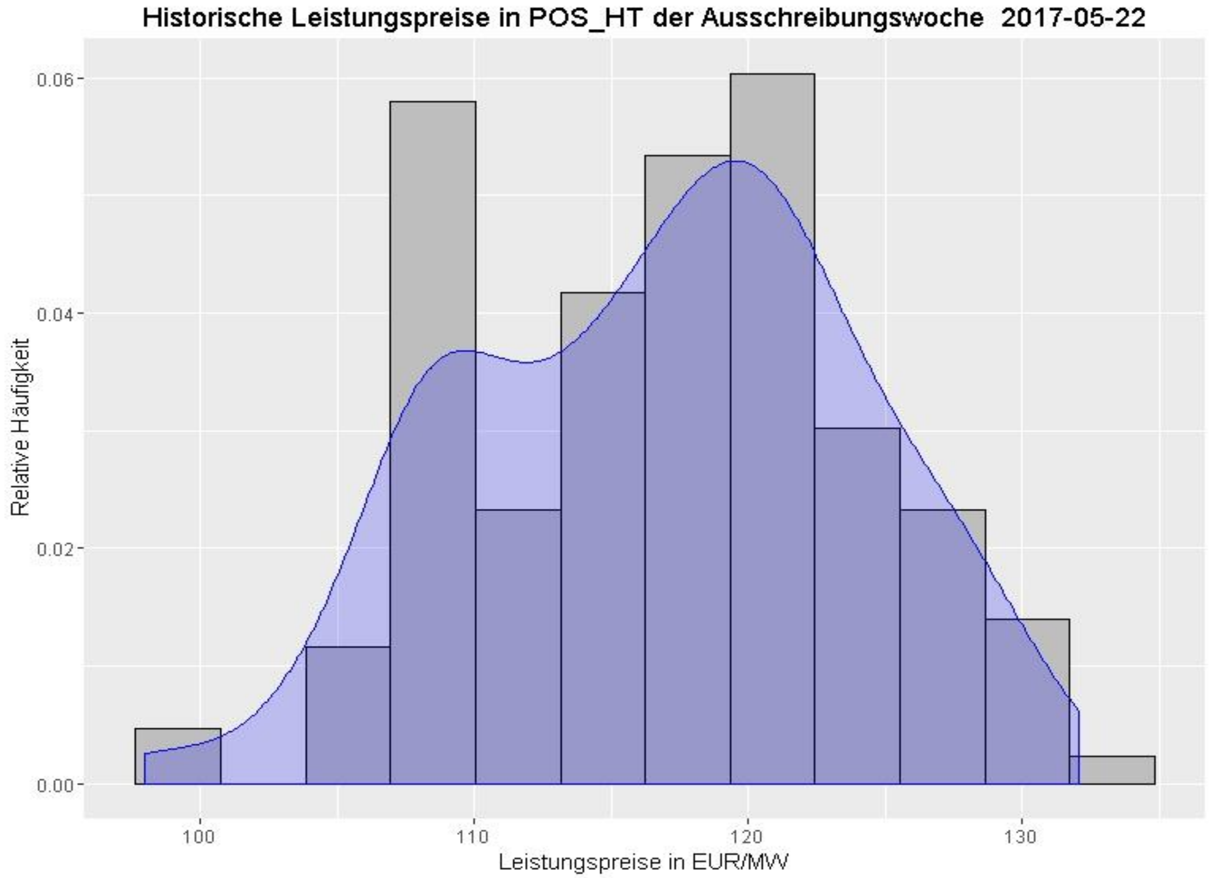
mit $\phi(y) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} y^2)$ Dichtefunktion von $N(0,1)$ und $\Phi(\cdot)$ die Verteilungsfunktion

- Numerisches Verfahren zur Parameterschätzung für μ und σ mittels der Maximum-Likelihood-Methode, d.h. Maximierung der Log-Likelihood-Funktion

$$L_T(\mu, \sigma | x_1, \dots, x_n) = -\frac{1}{2\sigma^2} \sum_{k=1}^{n} (x_k - \mu)^2 - n \log\left(\int_0^{GLP} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right) dy\right)$$

Fraunhofer ITWM

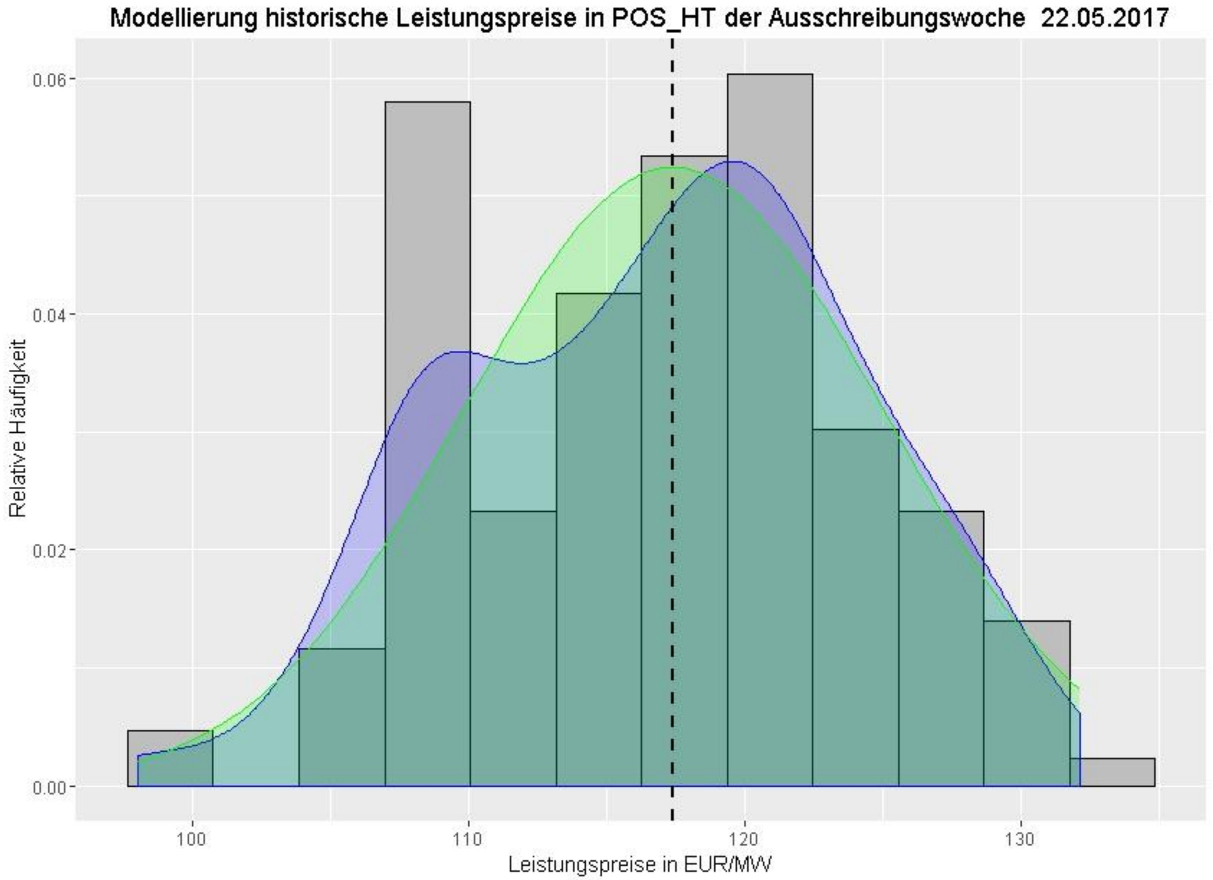
Die graphische Betrachtung der Woche mit Beginn am 22.05.17 zeigt, dass zwei Cluster in den Preisen existieren



- Cluster bei einem Leistungspreis von ca. 108 €/MW
- Cluster um den Leistungspreis von ca. 119 €/MW

Wie werden die Parameter der Normalverteilung geschätzt?

Die Modellierung mit der trunkierten Normalverteilung scheint bei Existenz von mehreren Clustern als ungeeignet



Resultate der Parameterschätzung mittels numerische Iterationsverfahren

$\mu = 117{,}4084$ und $\sigma^2 = 57{,}8983$

$\text{Leistungspreise} \sim N(117{,}4084,\ 57{,}8983)$

Beobachtungen zur Cluster-Modellierung

- Cluster in ca.108€/MW wird nicht abgebildet
- Aggregation um zweiten Preis von ca. 119€/MW wird leicht nach unten verschoben

Um die Heterogenität der Daten zu berücksichtigen, führen wir die Modellierung mit der Gausschen Mischverteilung ein

Bei der Gausschen Mischverteilung werden K Normalverteilungen verwendet – bei der Modellierung auf Basis abgeschnittener Daten muss die Dichte modifiziert werden:

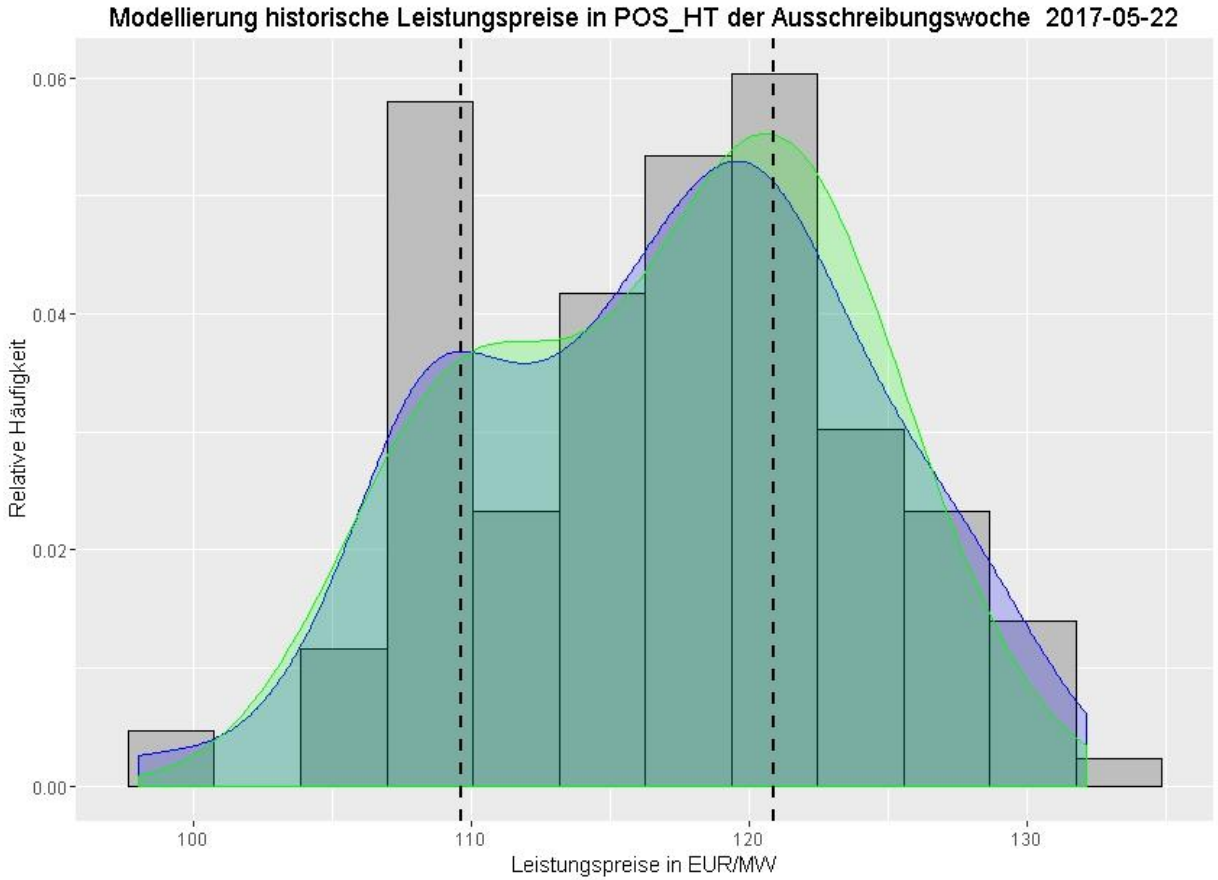
$$g(y|\Theta, t, s) = \sum_{k=1}^{K} \eta_k \, g_k(y^n|\mu_k, \sigma_k^2, t, s) \text{ mit}$$

$$g_k\left(y^n|\mu_k, \sigma_k^2, t, s\right) = \frac{\frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{\left(y^n - \mu_k\right)^2}{2\sigma_k^2}\right)}{\int_s^t \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{\left(y' - \mu_k\right)^2}{2\sigma_k^2}\right) dy'} \text{ die Dichte der trunkierten Normalverteilung und}$$

η_k Gewichtungsfaktor der k-ten Komponente für $k = 1, \ldots, K$ und $\sum_{k=1}^{K} \eta_k = 1$

Analog zur trunkierten Normalverteilung: Parameter müssen geschätzt werden - hier mittels Expectation-Maximization-Algorithmus

Die geschätzte Gaussche Mischverteilung mit K=2 berücksichtigt die Cluster aus der empirischen Verteilung in ihrer Parameterschätzung



Resultate der Parameterschätzung mittels numerische Iterationsverfahren

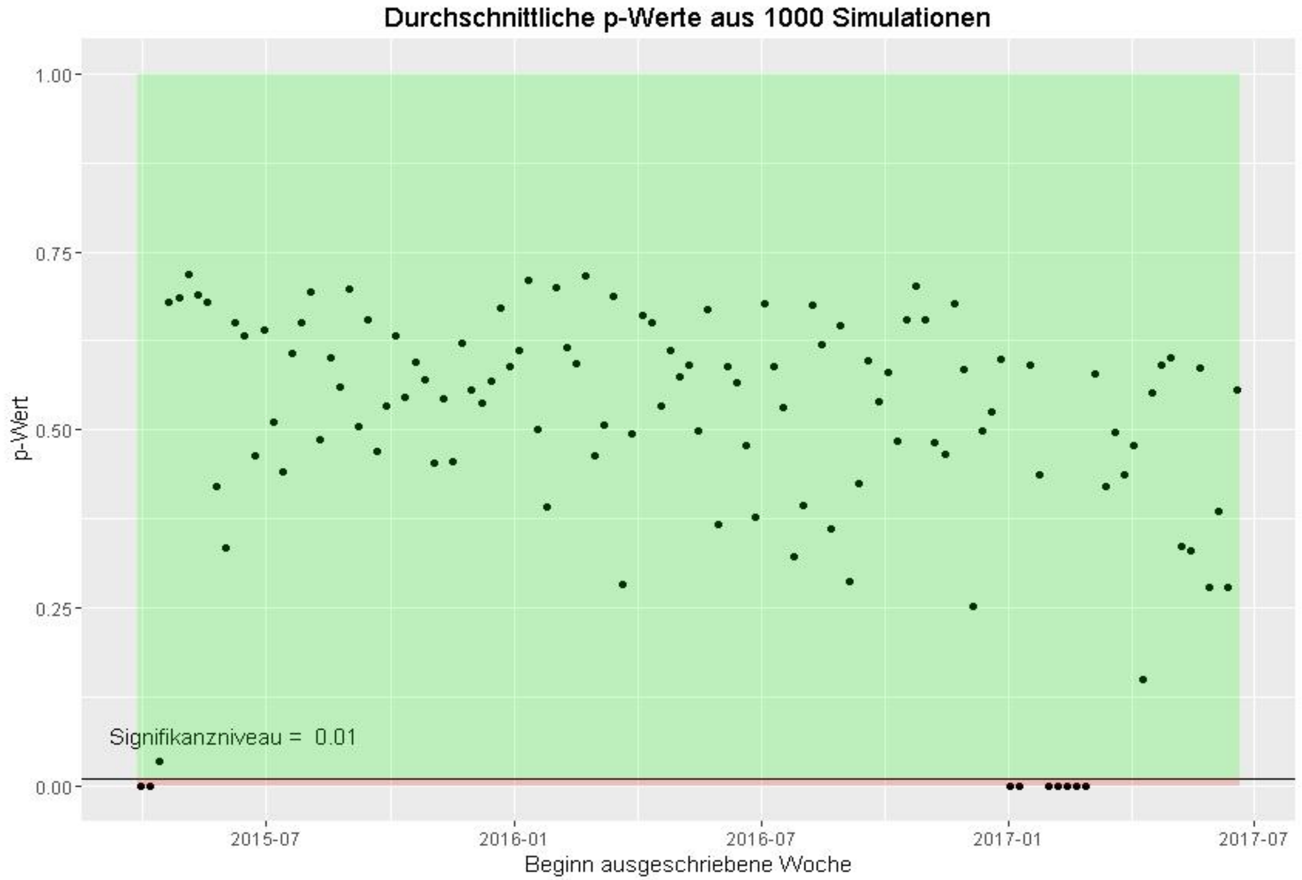
- $\mu_1 = 109,6064;\ \sigma_1^2 = 18{,}1382;\ \eta_1 = \ 0{,}3513$
- $\mu_2 = 120{,}8932;\ {\sigma_2}^2 = 22{,}8066;$

$$\eta_2 = (1 - 0{,}3513) = 0{,}6487$$

$$\text{Leistungspreise} \sim 0{,}3513\ N(109{,}6064; 18{,}1382) + 0{,}6487\ \mathrm{N}(120{,}8932; 22{,}8066)$$

Fraunhofer ITWM

Im Rahmen des Kolmogorov-Smirnov-Test mit simulierten Daten kann die Hypothese einer Gausschen Mischverteilung nicht verworfen werden



- Unsere Hypothesen:

  H_0: Leistungspreise folgen Gausschen-Mischverteilung

  H_1: Leistungspreise folgen keiner Gausschen-Mischverteilung
- Aufgrund unvollständiger Daten führen wir den Vergleich der empirischen Daten mit simulierten Daten aus der Verteilungsannahme
- Unsere Verteilungsannahme kann bei Modellierung von 116 Wochen in 107 Fällen nicht verworfen werden

Mit der 2-Komponenten-Gauss-Mischverteilung haben wir mehr Freiheitsgrade, um Clusterbildungen in den Leistungspreisen zu modellieren

Aspekte zur Modellierung der Leistungspreise und Ausblick

- Gaussche Mischverteilung scheint ein geeigneter Ansatz zu sein, um heterogene Daten zu modellieren – geht es vielleicht sogar besser mit $K > 2$?
- Modellierung der historischen Wochen als Grundlage zur Modellierung zukünftiger Wochen

Offene Diskussionspunkte:

- Betrachtung anderer Verteilungen zur besseren Modellierung?
- Grundannahme der Unabhängigkeit zwischen den Gebotskomponenten – Ist diese Annahme für die Realität plausibel?
- Die Unvollständigkeit der Daten zwingt zur Verwendung komplexerer Methoden – (Öffentlicher) Zugriff auf vollständige Datensätze evtl. vorhanden?

Fraunhofer ITWM

Literaturübersicht

- Cohen, A. (1949): **On Estimating The Mean and Standard Deviation of Truncated Normal Distributions, Journal of the American Statistical Association**, 45, 518–525
- Johnson, N. L., and Kotz, S. (1970): **Continuous Univariate Distributions**, Vol. 1 (Wiley Series in Probability and Statistics), Wiley-Interscience
- Lee G., Scott C. (2010): **EM algorithms for multivariate Gaussian mixture models with truncated and censored Data,** Journal Computational Statistics & Data Analysis 56 (2012)
- Biernackia C. , Celeuxb G. , Govaertc G. (2002): **Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models**, Journal Computational Statistics & Data Analysis 41 (2003)

Vielen Dank für Ihre Aufmerksamkeit!

Kontakt:
Prilly Oktoviany
Fraunhofer Platz 1
67663 Kaiserslautern
prilly.oktoviany@itwm.fraunhofer.de
0631-316004858

Einleitung und Definition des Sekundärregelleistungsmarktes

Derzeitiges Marktdesign des Sekundärregelleistungsmarktes

- **Wöchentliche** Ausschreibung mit Ausschreibungsbeginn mittwochs für die kommende Woche
- Gebotsabgaben für **vier Produkte**:
  - Positive SRL: **POS_HT** (08-20 Uhr wochentags) und POS_NT (20-08 Uhr wochentags und Wochenende ganztägig)
  - Negative SRL: NEG_HT und NEG_NT (gleiche Zeiteinteilung wie bei positiver SRL)
- Gebotsabgabe mit **drei Komponenten**: **Leistungspreis**, angebotene Leistung und Arbeitspreis
- Teilnahme an Ausschreibungen ausschließlich für präqualifizierte Teilnehmer
- Vor Ausschreibungsschluss: Geheime Gebotsabgabe auf regelleistung.net
- Ausschreibungsdesign: **Pay-As-Bid** nach **Merit-Order der Leistungspreise**
- Nach Ausschreibungsschluss: Veröffentlichung des Ausschreibungsergebnisses, jedoch nur von angenommenen Geboten, insbesondere keine Gebotsaufstellung über Grenzleistungspreis zu sehen

Die Parameterschätzung für Θ erfolgt mit dem Expectation-Maximization-Algorithmus

- **Ausgangspunkt:** Log-Likelihood-Funktion

$$L_T(\Theta) = \sum_n \sum_k z_k^n [ln(\eta_k) + ln(g_k(y^n))]$$

mit $\Theta = (\eta_1, \dots, \eta_K, \mu_1, \dots, \mu_K, \sigma_1, \dots, \sigma_K)$ zu schätzender Parametersatz, z_k^n Indikatorvariable

- **Idee EM-Algorithmus**: Gegeben die Daten und alte Parameterwerte, berechne im **E-Schritt**

$$Q_T\left(\Theta; \Theta^{\text{old}}\right) = \mathrm{E}[\mathrm{L_T}(\Theta) | \mathrm{y}^{1:\mathrm{N}}; \Theta^{old}]$$

den erwarteten Log-Likelihood und ermittle im

M-Schritt ein neues Θ^{new} sodass gilt

$$\Theta^{new} = \arg\max_{\Theta} Q_T(\Theta; \Theta^{old})$$

Fraunhofer
ITWM